Probability Comprehensive Exam Notes

Revised: December 31, 2018

Topics Index

1 Sigma algebras and sets

1.1 Set theory

Definition 1.1. We define the *limit infimum* and *limit supremum* of a sequence of sets by

$$\begin{aligned}
\liminf_{n\to\infty} A_n &= \bigcup_{n\geq 1}\bigcap_{m\geq n} A_m \\
&= \{x : x \text{ is in all but finitely many of the } A_n\text{'s}\} \\
\limsup_{n\to\infty} A_n &= \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m \\
&= \{x : x \text{ is in infinitely many of the } A_n\text{'s}\}.
\end{aligned}$$

Observe that for any $N \geq 1$:

$$\limsup_{n\to\infty} A_n \subseteq \bigcup_{i=N}^{\infty} A_i.$$

The following properties are also apparent:

- $\liminf_{n\to\infty} A_n = (\limsup_{n\to\infty} A_n^c)^c$;
- $\liminf(A_n) \subseteq \limsup(A_n)$.

Theorem 1.2 (Monotonicity of Sequences of Sets and Continuity)**.** *We say that A_n increases to A, written $A_n \nearrow A$, if $A_n \subset A_{n+1}$ for all $n \geq 1$ and $A = \cup_{n\geq 1} A_n$. We say that A_n decreases to A, written $A_n \searrow A$, if $A_n \supset A_{n+1}$ for all $n \geq 1$ and $A = \cap_{n\geq 1} A_n$. If $A_n \searrow A$, then by continuity we get that*

$$\lim_{n\to\infty} A_n = \bigcap_{n\geq 1} A_n.$$

Theorem 1.3 (Continuity of the Lebesgue measure on monotone sets from Real Analysis)**.** *Let $\{E_k\}$ be a sequence of measurable sets in $\mathbb{R}^n$.*

(1) If $E_k \nearrow E$, then $\lim_{k\to\infty} |E_k| = |E|$;

(2) If $E_k \searrow E$ and $|E_k| < \infty$ for some k, then $\lim_{k\to\infty} |E_k| = |E|$.

Proof of 1. If for some $j \geq 1$, $|E_j| = \infty$ then $|E_k| = \infty$ for all $k \geq j$, and hence $|E| = \infty$. Now we can assume that $|E_k| < \infty$ for all $k \geq 1$. We write

$$E = E_1 \cup (E_2 \setminus E_1) \cup \cdots = E_1 \cup \bigcup_{k\geq 1} (E_{k+1} \setminus E_k).$$

Then

$$|E| = |E_1| + \sum_{k\geq 1} |E_{k+1} \setminus E_k| = |E_1| + \lim_{n\to\infty} \sum_{k=1}^{n} (|E_{k+1}| - |E_k|) = \lim_{n\to\infty} |E_{n+1}|. \qquad \square$$

Proof of 2. We can assume that $|E_1| < \infty$ since otherwise we can truncate the sequence and relabel to obtain the same limit. Moreover, we can write E_1 as a disjoint union of measurable sets:

$$E_1 = E \cup \bigcup_{k\geq 1} (E_k \setminus E_{k+1}).$$

Now we have that

$$\begin{aligned}
|E_1| &= |E| + \sum_{k\geq 1} |E_k \setminus E_{k+1}| \\
&= |E| + \lim_{n\to\infty} \sum_{k=1}^{n} (|E_k| - |E_{k+1}|) \\
&= |E| + \lim_{n\to\infty} (|E_1| - |E_{n+1}|),
\end{aligned}$$

which implies the result. □

Proposition 1.4. *For X_n, Y_n random variables and any $a \in \mathbb{R}$, $\varepsilon > 0$, we have that*

$$\begin{aligned}
&\limsup\{Y_n > a+\varepsilon\} \subset \{\limsup Y_n > a\} \subset \{\limsup Y_n > a-\varepsilon\} \\
&\limsup\{Y_n < a-\varepsilon\} \subset \{\limsup Y_n < a\} \subset \{\limsup Y_n < a+\varepsilon\},
\end{aligned}$$

and that

$$\begin{aligned}
&\limsup_{n\to\infty}\{X_n > a+\varepsilon\} \subseteq \left\{\limsup_{n\to\infty} X_n > a\right\} \subseteq \limsup_{n\to\infty}\{X_n < a-\varepsilon\} \\
&\limsup_{n\to\infty}\{X_n < a-\varepsilon\} \subseteq \left\{\limsup_{n\to\infty} X_n < a\right\} \subseteq \limsup_{n\to\infty}\{X_n < a+\varepsilon\}.
\end{aligned}$$

1.2 Sigma algebras and sigma fields

We adopt the conventions that

$$\begin{aligned}
&\Omega \equiv \text{sample space} = \text{set of all possible outcomes}, \\
&\mathcal{F} \equiv \text{the collection of all events (subsets of the sample space)} \subseteq \rho(\Omega).
\end{aligned}$$

Definition 1.5 (Algebras and Sigma Fields)**.** The set (of sets) $\mathcal{F}$ is a *algebra* (or *field*) if

(1) $\Omega \in \mathcal{F}$;

(2) $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$;

(3) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.

Similarly, the set $\mathcal{F}$ is a σ-*algebra* (or σ-*field*) if

(1) $\Omega \in \mathcal{F}$;

(2) $A_n \in \mathcal{F} \implies \cup_{n\geq 1} A_n \in \mathcal{F}$;

(3) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.

Examples of σ-algebras include the following: $\mathcal{F}_1 = \{\varnothing, \Omega\}$ and $\mathcal{F}_2 = \rho(\Omega)$, the power set of Ω. We have the following consequences of these definitions:

- $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$ since $(A\cap B)^c = A^c \cup B^c$;
- $A, B \in \mathcal{F} \implies A \setminus B \in \mathcal{F}$;
- fields $\subseteq$ σ-fields;

- For $\mathcal{F}$ a σ-field, $A_n \in \mathcal{F} \implies \cap_{n\geq 1} A_n \in \mathcal{F}$.

Definition 1.6 (λ-Systems and π-Systems)**.** A set $G \subset \rho(\Omega)$ is called a λ-*system* if

(1) $\Omega \in G$;

(2) $A_n \in G$ (disjoint) $\implies \dot{\bigcup}_{n\geq 1} A_n \in G$;

(3) $A, B \in G$, $A \subset B \implies B \setminus A \in G$.

If C is a λ-system and in addition, C is closed under intersections (i.e., $A, B \in C \implies A \cap B \in C$), then C is called a π*-system*.

Definition 1.7 (Notation)**.** Let C be a collection of subsets in Ω. Then $\sigma(C)$ is defined to be the *smallest σ-algebra* containing C:

$$\sigma(C) = \bigcap_{\substack{C \subset \mathcal{F} \\ \mathcal{F} \text{ a } \sigma\text{-algebra}}} \mathcal{F}.$$

Similarly, we define $\lambda(C)$ to be the smallest λ-system containing C.

Theorem 1.8 (Dyukin)**.** *If C is a π-system, then $\lambda(C) = \sigma(C)$.*

Example 1.9 (The Borel Field of $\mathbb{R}$)**.** Consider the following sets:

$$\begin{aligned}
C_1 &= \{(a, b) : a < b\} \\
C_2 &= \{(a, b] : a < b\} \\
C_3 &= \{[a, b] : a < b\} \\
C_4 &= \{(-\infty, a] : a \in \mathbb{R}\} \\
C_5 &= \{[a, \infty] : a \in \mathbb{R}\} \\
C_6 &= \{[a, b] : a, b \in \mathbb{Q}\}.
\end{aligned}$$

Then the *Borel field* $\beta(\mathbb{R}) = \sigma(C_1) = \cdots = \sigma(C_6)$. It happens that all of the C_i's in this example are also examples of π-systems.

2 Probabilities

2.1 Definitions and basic properties

Throughout we will use the notation that Ω denotes the *sample space* and that $\mathcal{F}$ denotes the σ-field of events. Given a fixed probability $\mathbb{P}$, we will make reference to the *probability space* $(\omega, \mathcal{F}, \mathbb{P})$.

Definition 2.1. The function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a *probability* if

- $\mathbb{P}(\Omega) = 1$; and
- $(A_n)_{n\geq 1} \subset \mathcal{F}$ (disjoint) $\implies \mathbb{P}\left[\cup_{n\geq 1} A_n\right] = \sum_{n\geq 1} \mathbb{P}(A_n)$.

The next properties are apparent from the definition of P as a probability function:

- $\mathbb{P}(\varnothing) = 0$;
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, $\forall A \in \mathcal{F}$;
- $A \subset B \implies \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$, since $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.

Theorem 2.2 (Uniqueness on $\mathbb{R}$)**.** *If $\mathbb{P}, \mathbb{Q}$ are two probability measures on $\mathbb{R}$, and if $\mathbb{P}(-\infty, x] = \mathbb{Q}(-\infty, x]$ for all $x \in \mathbb{R}$, then $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \beta(\mathbb{R})$.*

Definition 2.3 (Continuity and Additivity)**.** A probability $\mathbb{P}$ is *continuous* if $A_n \searrow \varnothing \implies \mathbb{P}(A_n) \to 0$ as $n \to \infty$. The function $\mathbb{P}$ possessing *finite additivity* is equivalent to the property that whenever $(A_n)_{n\geq 1}$ are disjoint:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

On the other hand, σ-*additivity* is equivalent to

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Proposition 2.4. *If $\mathbb{P}$ is finite additive and continuous, then it is σ-additive.*

Example 2.5 (Properties of Probabilities)**.** We have the following additional special properties of probability function $\mathbb{P}$:

- **Subadditivity:** For any $(A_n) \subseteq \mathcal{F}$, $\mathbb{P}\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} \mathbb{P}(A_n)$;
- **Fatou:** If $(A_n) \subseteq \mathcal{F}$ is arbitrary, then

$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

- **Probabilities on countable sets:** If $\Omega = \{\omega_1, \omega_2, \ldots\}$ is countable, then for any $A \subseteq \Omega$: $\mathbb{P}(A) = \sum_{w_i \in A} \mathbb{P}(\omega_i)$.

2.2 Distribution functions (on $\mathbb{R}$ and $\mathbb{R}^n$)

Definition 2.6 (Distribution Functions)**.** If μ is a probability on $\beta(\mathbb{R})$, then the *distribution function of* μ is $F_\mu(x) := \mu((-\infty, x])$. We have the following properties of the distribution function:

(1) $\lim_{x\to-\infty} F_\mu(x) = 0$;

(2) $\lim_{x\to\infty} F_\mu(x) = 1$;

(3) F_μ is non-decreasing;

(4) F_μ is right continuous.

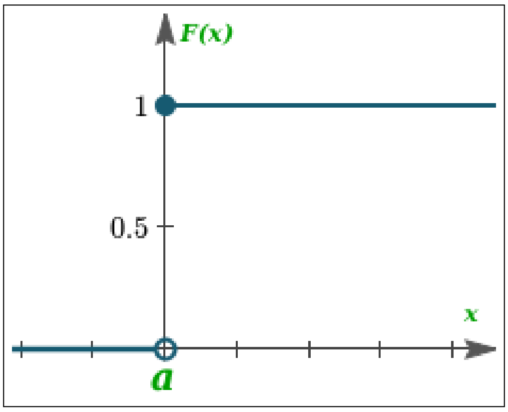
Theorem 2.7. *If u, v are probability measures on $\mathbb{R}$ such that $F_u(x) = F_v(x)$ for all $x \in \mathbb{R}$, then $u(A) = v(A)$ for all $A \in \beta(\mathbb{R})$.*

Theorem 2.8. *Let $(\Omega, \mathcal{F})$ be a probability space with $\mathcal{F} = \sigma(\mathcal{C})$ for $\mathcal{C}$ a π-system. If P, Q are two probabilities on $\mathcal{F}$ such that $P(C) = Q(C)$ for all $C \in \mathcal{C}$, then $P(A) = Q(A)$ for all $A \in \mathbb{F}$.*

Example 2.9 (The Dirac Distribution at a Point)**.** For a fixed $a \in \mathbb{R}$, we define

$$\mu(A) := \delta_a(A) = \begin{cases} 1, & a \in A; \\ 0, a \notin A. \end{cases}$$

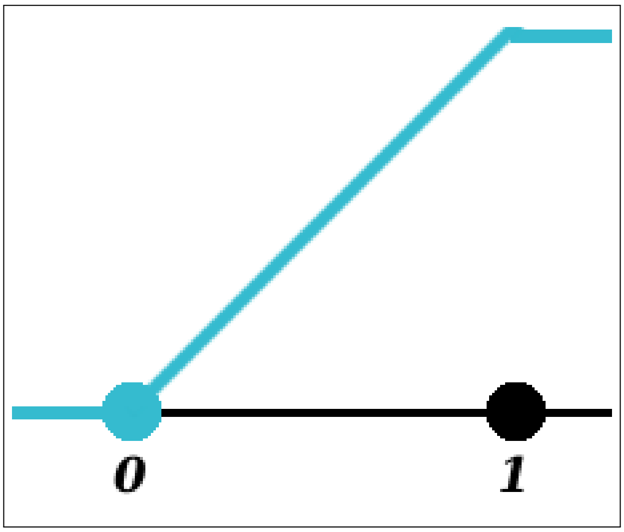
As we can check δ_a is a probability measure on $\mathbb{R}$ since $\delta_a(\mathbb{R}) = 1$ and (A_n) disjoint $\implies$ $\delta_a(\cup_{n\geq1} A_n) = \sum_{n\geq1} \delta_a(A_n)$. The distribution function of δ_a, F_{δ_a}, has the following plot shape for arbitrary $a \in \mathbb{R}$:



Example 2.10 (The Lebesgue Measure on the Unit Interval)**.** For an interval $[a, b] \subset [0, 1]$, we define the *Lebesgue measure* by the function $\lambda([a, b]) = b - a$ for $0 \leq a < b \leq 1$ where $\lambda([0, 1]^c) = 0$. Then by integration, we can see that the corresponding distribution function satisfies:

$$F_\lambda(x) = \begin{cases} 0, & x < 0; \\ x, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases}$$

This distribution function has the following plot:



Example 2.11 (Construction of μ, F_μ from a Function)**.** Provided that the function F satisfies the four required distrubution function properties from the definition above, we can construct a probability measure μ corresponding to this F using the next steps:

- Set $\mu((-\infty, x]) := F(x)$. Then $\mu((a,b]) = \mu((-\infty, b] \setminus (-\infty, a]) = F(b) - F(a)$ for $a < b$.
- To handle the closed intervals, take:

$$\mu([a,b]) = \lim_{n\to\infty} \mu((a - \frac{1}{n}, b]) = \lim_{n\to\infty} (F(b) - F(a - 1/n)) = F(b) - F(a^-).$$

- We can apply the *Caratheodory criterion* to see that if μ^* is an outer measure, then

$$\mathcal{A} := \{A \subset \mathbb{R} : \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A),\ \forall B \in \beta(\mathbb{R})\},$$

  is a σ-field.
- Finally, we can define $\mu := \mu^*\Big|_{\beta(\mathbb{R})}$.

Definition 2.12 (Probabilities on $\mathbb{R}^n$)**.** Here, for example, we will select $n := 2$ to illustrate the key properties. If μ is a probability on $\mathbb{R}^2$, then its distribution function $F_\mu(x,y) := \mu((-\infty, x] \times (-\infty, y])$ satisfies:

(1) $\lim_{x,y\to-\infty} F_\mu(x,y) = 0$;

(2) $\lim_{\substack{x\to\infty \\ y\to\infty}} F_\mu(x,y) = 1$;

(3) F_μ is non-decreasing in each variable;

(4) F_μ is right continuous in each variable.

Example 2.13 (Spring 2018, #7)**.** Let F_n, F be distribution functions such that $F_n \to F$ weakly. If F is continuois, show that

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \longrightarrow 0.$$

Example 2.14 (Fall 2018, #5)**.** For distribution functions F, G on the real line, define

$$L(F,G) := \inf \{\varepsilon > 0 : \forall t \in \mathbb{R}, F(t) \le G(t+\varepsilon) + \varepsilon, G(t) \le F(t+\varepsilon) + \varepsilon\}.$$

It is known that L is a metric. Prove that $L(F_n, F) \to 0$ as $n \to \infty$ if and only if F_n converges weakly to F.

2.3 Integration review and properties of distribution functions

The *mean value theorem* (MVT for derivatives) states that if f is defined and continous on $[a, b]$ and differentiable on (a, b), then $\exists c \in (a, b)$ such that $(b - a)f'(c) = f(b) - f(a)$. The *intermediate value theorem* (IVT) states that if f is continuous on $[a, b]$, then for all c in the range between $f(a)$ and $f(b)$, $\exists x \in (a, b)$ such that $f(x) = c$. The *MVT for integrals* states that if f is continuous on $[a, b]$ then $\exists c \in (a, b)$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f(t)dt.$$

Notice that f ***monotone increasing*** on $[a, b] \implies f \in \mathrm{BV}[a, b]$. If f has bounded variation on $[a, b]$, then we can express $f = g - h$ where g, h are both bounded and monotone increasing on $[a, b]$ (***Jordan decomposition theorem***). Note that these functions can be extended to be monotone increasing on all of $\mathbb{R}$ as well. Also, f monotone increasing implies that f is differentiable a.e., $f' \geq 0$, and $\int_a^b f'(t)dt \leq f(b) - f(a)$ for all $a < b$.

Proposition 2.15 (Monotone increasing function properties)**.** *If f is monotone increasing, then*

(a) f has at most countably many points of discontinuity.

(b) f is differentiable a.e.

(c) f' is a measurable function.

(d) $f' \in L^1$ and we have the FTC as an upper bound: $0 \leq \int_a^b f' \leq f(b) - f(a)$.

Nested characterizations: We say that f is *Lipschitz on* $[a, b]$ if there exists a $K > 0$ (where K is independent of all $x, y \in [a, b]$) such that $|f(x) - f(y)| \leq K \cdot |x - y|$ for all $x, y \in [a, b]$. We have the following important nested inclusion of function types (C^1 denotes the set of functions with one continuous derivative):

$$C^1[a, b] \subsetneq \mathrm{Lip}[a, b] \subsetneq \mathrm{AC}[a, b] \subsetneq \mathrm{BV}[a, b] \subsetneq L^\infty[a, b].$$

Typical counter examples to these types of functions are $\sqrt{x}$ or are of the form $f(x) = x^a \sin(1/x^b)$.

Example 2.16 (Analysis Exam, Fall 2008, #3)**.** Let f, g be absolutely continuous functions on $[0, 1]$. Show that for $x \in [0, 1]$ we have

$$\int_0^x f(t)g'(t)dt = f(x)g(x) - f(0)g(0) - \int_0^x f'(t)g(t)dt.$$

(HINT: Consider the integral $\int\int_E f'(t)g'(t)dt$ over the set $E := \{(s, t) \in [0, x]^2 : s \leq t\}$.)

2.4 The product construction

Definition 2.17 (Products)**.** Suppose that $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_1, \mathcal{F}_2, P_2)$ are two probability spaces. Then the *product construction* between these two spaces corresponds to taking $\Omega := \Omega_1 \times \Omega_2$, $\mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\})$, and the tensor of the probabilities as $\mathbb{P} := P_1 \otimes P_2$, where $\mathbb{P}(A \times B) = P_1(A) \cdot P_2(B)$.

3 Independence

3.1 Basics and preliminaries

For our probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in general for any $A, B \in \mathcal{F}$, the conditional probability of A given B has occurred is

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

As a notion of *independence* of the random variables A and B, we want that observing B does not affect the probability of A, i.e., that $\mathbb{P}[A|B] = \mathbb{P}[A]$ provided that $\mathbb{P}[B] \neq 0$. This is equivalent to the condition that $\mathbb{P}[A|B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$, which holds even when $\mathbb{P}[B] = 0$. Thus we take this property as our definition of independence.

Example 3.1. First, if $A := \varnothing, \Omega$, then A, B are always independent for any $B \in \mathcal{F}$. We have that if A, B are independent, then A^C, B are independent. We can prove this fact using the following logic:

$$\begin{aligned}
\mathbb{P}(A^C \cap B) &= \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B), \quad \text{by independence,} \\
&= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A))\mathbb{P}(B) \\
&= \mathbb{P}(A^C)\mathbb{P}(B).
\end{aligned}$$

So to summarize, any of the elements of $\{\varnothing, \Omega, A, A^C\}$ are independent of any in $\{\varnothing, \Omega, B, B^C\}$ whenever A, B are independent.

We generalize to the independence of a set of random variables by defining that $A_1, A_2, \ldots, A_n$ are independent if $\forall 1 \leq i_1 < i_2 < \cdots < i_k \leq n$ (with $2 \leq k \leq n$) $\implies$

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

Note that this definition implies pairwise independence, but is actually a much stronger requirement. For example, A, B, C are independent if each of $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. In general, for I some indexing set, we make the next definition: $(\mathcal{F}_i)_{i \in I}$ are independent if $\forall i_1, \ldots, i_n \in I$ (distinct) with $A_{i_j} \in \mathcal{F}_{i_j}$ we have

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_n}) = \prod_{j=1}^{n} \mathbb{P}\left(A_{i_j}\right).$$

Theorem 3.2 (Generation Theorem)**.** *For any $i \in I$, assume that $\mathcal{F}_i \in \sigma(C_i)$ for C_i always a π-system. Then $(\mathcal{F}_i)_{i \in I}$ are independent if and only if $\forall i_1, \ldots, i_n \in I$ (distinct), and with $A_{i_j} \in C_{i_j}$, then $\mathbb{P}(C_{i_1} \cap \cdots \cap C_{i_n}) = \mathbb{P}(C_{i_1}) \cdots \mathbb{P}(C_{i_n})$.*

Corollary 3.3 (Grouping Theorem)**.** *If $(\mathbb{F}_i)_{i \in I}$ are independent, and if $\{I_j\}_{j \in J}$ partition I as $I = \dot{\cup}_{j \in J} I_j$, and if $\sigma_j := \sigma(\{\mathcal{F}_i : i \in I_j\})$, then the $(\sigma_j)_{j \in J}$ are independent.*

Example 3.4 (Independent Groupings)**.** Suppose that the sequence of $(A_n)_{n \geq 1}$ are independent. Then for

$$\begin{aligned}
B_1 &:= A_1 \cup A_3 \cup A_5 \cup A_7 \cup \cdots \\
B_2 &:= (A_2 \cap A_4 \cap A_6) \cup (A_8 \cap A_{10} \cap A_{12}) \cup \cdots \\
C_1 &:= A_1 \cup A_4 \cup A_7 \cup A_{10} \cup \cdots \\
C_2 &:= (A_2 \cap A_5) \cup (A_8 \setminus A_{11}) \cup \cdots \\
C_3 &:= (A_3 \setminus A_6) \cup (A_9 \setminus A_{12}) \cup \cdots,
\end{aligned}$$

we have that B_1, B_2 and C_1, C_2, C_3 are each independent from the "grouping" corollary above.

Example 3.5 (*A* Independent of Itself)**.** Observe that A is independent of itself (henceforth, a.s. constant per exam 2), then $\mathbb{P}(A) \in \{0,1\}$. This happens because of our independence requirement that $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2 \implies \mathbb{P}(A) \in \{0,1\}$.

Example 3.6 (A Three-Variable Proof)**.** If A, B, C are independent, then we can prove also that A^C, B, C are independent:

$$\begin{aligned}\mathbb{P}\left(A^C \cap B \cap C\right) &= \mathbb{P}\left((B \cap C) \setminus A\right) = \mathbb{P}(B \cap C) - \mathbb{P}(B \cap C \cap A) \\ &= \mathbb{P}(B)\,\mathbb{P}(C) - \mathbb{P}(B)\,\mathbb{P}(A)\,\mathbb{P}(C) = (1 - \mathbb{P}(A))\mathbb{P}(B)\,\mathbb{P}(C) \\ &= \mathbb{P}\left(A^C\right)\mathbb{P}(B)\,\mathbb{P}(C)\,.\end{aligned}$$

Example 3.7 (Fall 2016, #6)**.** Let $\{A_n\}$ be an infinite collection of independent events. Suppose that $\mathbb{P}(A_n) < 1$ for every $n \geq 1$. Show that $\mathbb{P}(A_n, i.o.) = 1$ if and only if $\mathbb{P}(\cup A_n) = 1$.

3.2 Key examples and results

Theorem 3.8 (Kolmogorov's $0-1$ Law)**.** *Let* $*\mathcal{F}_n)_{n\geq 1}$ *be an infinite sequence of* σ*-fields. The corresponding, so-called tail field is*

$$T := \bigcap_{n\geq 1} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \ldots),$$

where the sets $\sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \ldots)$ *are decreasing with* n*. If the* $(\mathcal{F}_n)_{n\geq 1}$ *are independent, then* $\mathbb{P}(A) \in \{0,1\}$ *for all* $A \in T$*. As a consequence of this phenomenon, we observe that if* $(A_n)_{n\geq 1}$ *are independent, then*

(1) $\mathbb{P}(\limsup_{n\to\infty} A_n) \in \{0,1\}$*; and*

(2) $\mathbb{P}(\liminf_{n\to\infty} A_n) \in \{0,1\}$.

Theorem 3.9 (Borel-Cantelli Lemma)**.** *If* $(A_n)_{n\geq 1}$ *is such that* $\sum_{n\geq 1}\mathbb{P}(A_n) < \infty$*, then importantly,* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$*. This observation is later employed to guarantee almost sure convergence of the* A_n *to zero provided that the latter sum converges. Secondly, if* $(A_n)_{n\geq 1}$ *are independent, then*

$$\sum_{n\geq 1} \mathbb{P}(A_n) = +\infty \implies \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

NOTE: We can utilize the trick that for any set $A \in \mathbb{F}$, $\mathbb{P}\left(A^C\right) = 1 - \mathbb{P}(A) \leq e^{-\mathbb{P}(A)}$, which is essentially the same as saying that $1 - x \leq e^{-x}$ whenever $x \in [0,1]$.

Example 3.10 (Spring 2018, #8)**.** Let $(X_n)_{n\geq 1}$ be an iid sequence of random variables. Show that $E[X_1^2] < \infty$ if and only if for every $c > 0$, $\mathbb{P}(|X_n| \geq c\sqrt{n}$ infinitely often$) = 0$.

Example 3.11 (Preview of Independent Bernoulli Random Variables)**.** For $n \geq 1$, let $X_n \sim$ Bernoulli(p_n) so that $\mathbb{P}(X_n = 1) = p_n$ and $\mathbb{P}(X_n = 0) = 1 - p_n$. Then we claim that $X_n \xrightarrow{a.s.} 0 \iff \sum_{n\geq 1} p_n < \infty$. This is true since if the right-hand-side sum converges then

$$\sum_{n\geq 1} \mathbb{P}(X_n = 1) = \sum_{n\geq 1} \mathbb{P}(|X_n - 0| > \varepsilon) < \infty,$$

so that by Borel-Cantelli $\limsup_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0$.

Example 3.12. If $X_1, \ldots, X_n$ are independent, then for $\phi_1, \ldots, \phi_n : \mathbb{R} \to \mathbb{R}$ measurable, the random variables $Y_i := \phi_i(X_i)$ are still independent. Why is this the case?

$$\{Y_i \in A_i\} = \{\phi_i(X_i) \in A_i\} = \{X_i \in \phi_i^{-1}(A_i)\},$$

where the $\phi_i^{-1}(A_i)$'s are measurable. Hence, the Y_i's are independent as well.

4 Random variables

4.1 Definitions and basic properties

Definition 4.1. A function $X : \Omega \to \mathbb{R}$ defined on the sample space is a *random variable* if $X^{-1}(A) \in \mathcal{F}$ for all $A \in \beta(\mathbb{R})$. Here, we define $X^{-1}(A) := \{\omega : X(\omega) \in A\} = \{X \in A\}$ where the second set notation is used more "loosely" as a simplification to use as a nice tool for understanding. We immediately get the following properties which are apparent given our nice new notation for expressing the meaning of a r.v. X:

(1) $X^{-1}\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i \in I} X^{-1}(A_i)$;

(2) $\left\{X \in \bigcap_{i\in I} A_i\right\} = \bigcap_{i\in I}\{X \in A_i\}$; and

(3) $\{X \in A^C\} = \{X \in A\}^C$, or in other words, $X^{-1}(A^C) = X^{-1}(A)^C$.

For any fixed random variable X, we get a corresponding *probability measure* μ_X defined by

$$\mu_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(X^{-1}(B)\right),\ \forall B \in \beta(\mathbb{R}).$$

Generalizations: We can extend this notion more generally by setting $(\Omega, \mathcal{F}) \mapsto (S, \varsigma)$, with $\mathcal{F}, \varsigma$ both σ-fields. In particular, a S-valued random variable is a function $X : \Omega \to S$ such that $\{X \in A\} \in \mathcal{F}$ for all $A \in \varsigma$. The next couple of results fill in some key details for extended random variables of this type.

Proposition 4.2. *If $X : \Omega \to S$ is a random variable, and ς a σ-field, then $\mathcal{F}_X = X^{-1}(\varsigma) = \{X^{-1}(A) : A \in \varsigma\}$ is a σ-field. Moreover, if $X : \Omega \to S$, with $\varsigma = \sigma(\mathcal{C}')$, then $\sigma(X^{-1}(\mathcal{C}')) = X^{-1}(\varsigma) = X^{-1}(\sigma(\mathcal{C}'))$.*

Corollary 4.3 (Characterization for the $\mathbb{R}$ Case)**.** *If $S = \sigma(\mathcal{C}')$, then $X : \Omega \to S$ is a random variable if and only if $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{C}'$. In the real-field case, $X : \Omega \to \mathbb{R}$ is a random variable if and only if $\forall x \in \mathbb{R}$, the sets $\{X \leq x\} \in \mathcal{F}$.*

Proposition 4.4 (Useful General Facts for the Real Case)**.** *We have the next several observations which are collected together here as a matter of reference and completeness:*

(1) If $X_1, \ldots, X_n$ are random variables on Ω, then the tuple $X \equiv (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ is a $\mathbb{R}^n$-valued random variable;

(2) If $g : \mathbb{R}^n \to \mathbb{R}$ is a measurable function, i.e., if $g^{-1}(A) \in \beta(\mathbb{R}^n)$ for all $A \in \beta(\mathbb{R})$, then $g(X_1, \ldots, X_n)$ is a random variable when the X_i's are;

(3) If $g : \mathbb{R}^n \to \mathbb{R}$ is continuous, then g is measurable;

(4) Examples of measurable functions on $\mathbb{R}^2$ include: $g(x, y) = x + y, x - y, xy, \frac{x}{y}$ $(y \neq 0)$, and $g(x, y) = \chi_A(x)\chi_B(y)$ for $A, B \in \beta(\mathbb{R})$ are all measurable, let's say where they are defined.

Example 4.5 (Limits)**.** If $\{X_n\}_{n\geq 1}$ is any sequence of random variables, then

$$Y := \inf X_n, \sup X_n, \liminf X_n, \limsup X_n,$$

are all random variables.

Example 4.6 (Spring 2017, # 4)**.** Let X, Y be independent, and suppose that each has a uniform distribution on $(0, 1)$. Let $Z := \min(X, Y)$. Find the density $f_Z(z)$ for Z.

4.2 Independence

Definition 4.7 (Independence)**.** If $(X_i)_{i\in I}$ are random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P}\left(\right))$, we say that the sequence is independent if $(\mathcal{F}_{x_i})_{i\in I}$ are independent. Here, we write $\mathcal{F}_{x_i} = \{X_i^{-1}(A) : A \in \beta(\mathbb{R})\}$.

The next is a key characterization concerning the independence of random variables: $(X_i)_{i\in I}$ are independent if whenever $i_1, \dots, i_n \in I$ are distince,

$$\mathbb{P}\left(X_{i_1} \leq x_{i_1}, \dots, X_{i_n} \leq x_{i_n}\right) = \mathbb{P}\left(X_{i_1} \leq x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_n} \leq x_{i_n}\right),$$

for any $x_{i_1}, \dots, x_{i_n} \in \mathbb{R}$.

Proposition 4.8. *If $X_1, \dots, X_n$ are independent random variables and $g_1, \dots, g_n$ are Borel measurable functions, then $g_1(X_1), \dots, g_n(X_n)$ are also independent.*

Proposition 4.9. *If X, Y are independent random variables with both $E[X], E[Y] < \infty$, then $E[|XY|] < \infty$ and $E[XY] = E[X]E[Y]$.*

Remark 4.10 (Orthogonal Random Variables)**.** Random variables X, Y such that $E[XY] = E[X]E[Y]$ are called *orthogonal*. It is clear from the above that independent random variables with finite expectation are orthogonal. We also have that if $X_1, \dots, X_n$ are random variables with finite variance, and if $X_1, \dots, X_n$ are pairwise orthogonal, then

$$\operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n).$$

Example 4.11 (Spring 2018, #3)**.** Let X, T be random variables with $E[X], E[Y] < \infty$. Prove that if $E[X|Y] = Y$ and $E[Y|X] = X$, then $X = Y$ a.s.

Example 4.12 (Fall 2018, #2)**.** Suppose $X_1, \dots, X_n$ are iid random variables such that $\mathbb{P}\left(X_j = +1\right) = \mathbb{P}\left(X_j = -1\right) = 1/2$. Let $S_k := X_1 + \cdots + X_k$ for $1 \leq k \leq n$. Prove that

$$\mathbb{P}\left(\max_{1\leq k\leq n} S_k \leq \ell\right) = 2\mathbb{P}\left(S_n > \ell\right) + \mathbb{P}\left(S_n = \ell\right).$$

Conditional expectation

Example 4.13 (Fall 2018, #8)**.** Let X and Y be two independent and positive random variables with respective density F_X, f_Y, and let $g : (0, \infty) \to (0, \infty)$ be a bounded Borel function. Find

$$E\left[g\left(\frac{X}{Y}\right)\Big|Y\right],$$

the conditional expectation of $g(X/Y)$ given Y, and then infer that $V := X/Y$ has a density that you will identify.

Example 4.14 (Fall 2018, #9)**.** Let X, Y, Z be random variables such that (X, Z) and (Y, Z) are identically distributed. Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function such that $f(X)$ is integrable.

(i) Show that $E[f(X)|Z] = E[f(Y)|Z]$, a.s.; and

(ii) Let $T_1, \dots, T_n$ be iid random variables with finite first moment, and let $T = T_1 + \cdots + T_n$. Using (i), show that

$$E[T_1|T] = \frac{T}{n}.$$

4.3 Common distributions

Definition 4.15. We introduce models for some of the most common discrete probability distributions below. In each of the following cases, $X : \Omega \to \mathbb{R}$ (or possibly $\mathbb{R}^n$) is a random variable, and it's probability distribution, or measure function which is what varies from case to case, is denoted by μ_X:

- **Bernoulli:** For a fixed probability $p \in (0,1)$, we say that $X \sim \text{Bernoulli}(p)$ if $\mathbb{P}(X=1) = p$ and $\mathbb{P}(X=0) = 1-p$. Here, we express the measure function by $\mu_X = p\delta_1 + (1-p)\delta_0$, where
$$\delta_a(A) = \begin{cases} 1, & a \in A; \\ 0, & a \notin A, \end{cases}$$
is the *Dirac delta function*. Indeed, we can see that this expansion is correct since 1) μ_X is additive (because the δ_i's are); and 2) it is a probability measure as $\mu_X(\mathbb{R}) = p\delta_1(\mathbb{R}) + (1-p)\delta_0(\mathbb{R}) = 1$. It is standard information that $E[X] = p$, and $\text{Var}(X) = pq$, where $q = 1-p$ is the flipped probability for the distribution. Note that a sum of n iid Bernoulli random variables with parameter p have mean $\mu = np$ and variance $np(1-p)$.

- **Uniform:** $X \sim$ Uniform (on a finite, or countable set) if $S = \{x_1, \ldots, x_n\}$ and we have that $\mu_X(\{x_1\}) = \cdots = \mu_X(\{x_n\}) = \frac{1}{n}$. Alternately, we can state this condition as $X : \Omega \to \mathbb{R}$ is uniform on S provided that $\mathbb{P}(X = x_1) = \cdots = \mathbb{P}(X = x_n) = \frac{1}{n}$.

- **Geometric:** $X \sim \text{Geom}(p)$, which models
$$X := \#\text{ flips in a sequence of coin tosses until the first } H \text{ occurs.}$$
We have: $\mathbb{P}(X = k) = p(1-p)^{k-1}$ for any natural numbers $k \geq 1$.

- **Binomial:** $X \sim \text{Binomial}(n,p)$, where X models $X := \#$ H's in n coin flips, and is such that $\mathbb{P}(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$ for $0 \leq k \leq n$. We can show that $E[X] =: \mu = np$, and $\text{Var}(X) = np(1-p)$.

- **Poisson:** $X \sim \text{Poisson}(\lambda)$, for $\lambda > 0$, if for all integers $k \geq 0$: $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. The *Poisson distribution* naturally arises as the limit of a binomial distribution when the expected number of successes stays fixed at λ: namely, the limiting distribution of $X \sim \text{Binomial}(n, \lambda/n)$ as $n \to \infty$. It can be verified that if X is Poisson with parameter $\lambda > 0$, then $E[X] = \lambda$, $E[X(X-1)] = \lambda^2$, and $\text{Var}(X) = E[X^2] - E[X]^2 = \lambda$.

Example 4.16 (Preview of Continuous Random Variable Distributions)**.** X is a *continuous random variable* if its probability measure, μ_X, has a consistent density function. Namely, f satisfying $f \geq 0$, $\int_{\mathbb{R}} f = 1$, and where $\mu_X(A) = \int_A f(x)dx$. The following listing provides a brief overview of the properties of several of the most common cases we will see:

- **Exponential:** With the construction of this model distribution, we have that the density function is non-zero, and equal to $f(x) = \lambda e^{-\lambda x}$, iff and only if $x > 0$. It is defined to be zero otherwise. Here, to ensure that we get a resulting probability measure, the requirement is that $\lambda > 0$, for example as in other instances in the next sections, with $\lambda := 1$. Note that since we have that $\mathbb{P}(X \geq t) = e^{-\lambda t}$, $\forall t \geq 0$, we can compute that
$$E[X] = \int_0^\infty \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}$$
$$E[X^2] = \frac{2}{\lambda^2}$$

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{1}{\lambda^2}.$$

The exponential distribution has the so-called *loss-of-memory property*, which is to say that for all $s, t \geq 0$, $\mathbb{P}(X \geq s + t | X \geq s) = \mathbb{P}(X \geq t)$.

- **Uniform:** A.k.a., the non-discretized version of a uniformly distributed random variable on some $[a, b] \subset \mathbb{R}$. Then the density and cumulative density functions respectively have the formulas:

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & \text{otherwise}, \end{cases}$$

$$F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & x \in [a, b); \\ 1, & x \geq b. \end{cases}$$

Here, we have that $\mu = \frac{1}{2}(a + b)$ and $\sigma^2 = \frac{(b-a)^2}{12}$, where the density function can be rewritten as

$$f(x) = \begin{cases} \frac{1}{2\sigma\sqrt{3}}, & -\sigma\sqrt{3} \leq x - \mu \leq \sigma\sqrt{3} \\ 0, & \text{otherwise}. \end{cases}$$

- **Standard Normal:** $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. See Section 7 for a substantially complete treatment. Note that $X_i \sim N(0, 1) \implies Y := \sum_{i=1}^n \alpha_i X_i \sim N(0, \alpha_1^2 + \cdots + \alpha_n^2)$. If $X_i \sim N(\mu_j, \sigma_j^2)$, then $Y := \sum_{i=1}^n X_i \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots \sigma_n^2)$. If $X_i \sim N(0, 1)$, then $Z_n = X_1 + \cdots + X_n$ has the same distribution as $\sqrt{n}Z$.

- **Gamma:** For parameters $\lambda, \alpha > 0$, we say that $X \sim \mathrm{Gamma}(\lambda, \alpha)$ if $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ for $x > 0$. Here, the *gamma function* is defined by the usual integral $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. We have the following properties of compositite (sum) distributions which result in gamma-distributed random variables:

  (G.0) If $\alpha = 1$ then $\mathrm{Gamma}(\lambda, 1) \equiv \mathrm{Exp}(\lambda)$ is the exponential distribution;

  (G.1) If $X_1, \ldots, X_n$ are iid with $X_1 \sim \mathrm{Exp}(\lambda)$, then $X_1 + \cdots + X_n \sim \mathrm{Gamma}(\lambda, n)$;

  (G.2) More generally, if $X_1, \ldots, X_n$ are iid random variables with $X_i \sim \mathrm{Gamma}(\lambda, \alpha_i)$, then $X_1 + \cdots + X_n \sim \mathrm{Gamma}(\lambda, \alpha_1 + \cdots + \alpha_n)$'

  (G.3) If $Z \sim N(0, 1)$ is standard normal, then $Z^2 \sim \mathrm{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$;

  (G.4) If $n \geq 1$ is an integer then the distribution $\mathrm{Gamma}\left(\frac{1}{2}, \frac{n}{2}\right)$ is called the χ^2*-distribution with* n *degrees of freedom*, which also corresponds to the distribution of $Z_1^2 + \cdots + Z_n^2$ where $Z_1, \ldots, Z_n$ are iid standard normal random variables.

Exercise 4.17. Show that if $X \sim \mathrm{Exp}(\lambda)$ and $U := e^{-X}$ that $U \sim \mathrm{Uniform}(0, 1)$.

Example 4.18 (Spring 2017, #8)**.** Assume $X_1, X_2, \ldots$ are iid standard normal random variables. Show that for any $\lambda > 1/2$,

$$\frac{1}{n^\lambda}(X_1 + \cdots + X_n) \xrightarrow{a.s.} 0.$$

Example 4.19 (Fall 2016, #2)**.** Let X be a random variable with continuous density function f and $f(0) > 0$. Let Y be a random variable with

$$Y = \begin{cases} \frac{1}{X}, & \text{if } X > 0; \\ 0, & \text{otherwise}, \end{cases}$$

and let $Y_1, Y_2, \ldots$ be iid with distribution equal to that of Y. What is the value of the almost sure limit:
$$\lim_{n\to\infty} \frac{Y_1 + \cdots + Y_n}{n}?$$

Example 4.20 (Spring 2016, #3)**.** Let $X_1, X_2, \ldots$ be iid uniform random variables on $(0,1)$. Show that
$$(X_1 \cdots X_n)^{1/n},$$
converges almost surely as $n \to \infty$, and compute this limit.

4.4 Expectation

Definition 4.21. If $X = \chi_A$, then $E[X] = \mathbb{P}(A)$. We also use the notation $E[X, A] := E[X\chi_A]$. A random variable X is called *simple* if $X = \sum_{i=1}^N a_i \chi_{A_i}$ where the A_i's are disjoint. This implies that $E[X] = \sum_{i=1}^N a_i \mathbb{P}(A_i)$. If $X \geq 0$, then $E[X] = \int X d\mathbb{P}$. If X is an arbitrary random variable, then $X = X_+ - X_-$ where $X_+ = \max(X, 0)$ and $X_- = \max(-X, 0)$. Here we have that $|X| = X_+ + X_-$. We say that X is *integrable* if $E[X_+] < \infty$ and $E[X_-] < \infty$.

We make the following observations:

- Note that for X simple, $E[X]$ does not depend on the representation of X. That is, if $X = \sum_i a_i \chi_{A_i} = \sum_j b_j \chi_{B_j}$, then $\sum_i a_i \mathbb{P}(A_i) = \sum_j b_j \mathbb{P}(B_j)$.
- If $X \geq 0$, $\exists$ a sequence $(X_n)_{n\geq 1}$ of simple random variables such that $X_n \nearrow X$ and
$$\lim_{n\to\infty} E[X_n] = E[X],$$
where this limit is possibly infinite.
- To define a general random variable $X \geq 0$ from simple functions, use a limiting procedure: $X_n := \sum_{k=1}^{n^2} \frac{(k-1)}{n} \chi_{\frac{k-1}{n} \leq X \leq \frac{k}{n}}$, so that $X_n \nearrow X$ and $\lim_{n\to\infty} X_n = X$.

Expectation of a Discrete Random Variable: Suppose that $X : \Omega \to \{a_1, a_2, a_3, \ldots\}$ is a *discrete random variable.* Then
$$E[X] = \sum_{j\geq 1} a_j \cdot \mathbb{P}(X = a_j).$$

Proposition 4.22. *The random variable is integrable* $\iff$ $E[|X|] < \infty$ *(i.e., X is integrable if and only if $|X|$ is integrable).*

Proposition 4.23 (Properties of Expectation)**.** *We have the following key properties:*

(1) $E[X] \geq 0$ *if* $X \geq 0$.

(2) If $X \geq 0$, *then* $E[X] = 0 \iff X = 0$ *almost surely.*

(3) $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$.

(4) (Monotone convergence theorem) If $X_n \geq 0$ *and* $X_n \nearrow X$, *then* $E[X_n] \nearrow E[X]$

(5) $|E[X]| \leq E[|X|]$.

(6) If X, Y are independent and integrable, then XY is integrable, and $E[XY] = E[X]E[Y]$.

(7) *If X, Y are such that for all bounded, smooth functions $\phi, \psi : \mathbb{R} \to \mathbb{R}$, $E[\phi(X)\psi(Y)] = E[\phi(X)]E[\psi(Y)]$, then X, Y are independent.*

(8) *(Fatou's Lemma) If $X_n \geq 0$, then $E\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} E[X_n]$.*

(9) *(Dominated Convergence Theorem) If $|X_n| \leq Y$ for all $n \geq 1$ with Y integrable, and if $X_n \xrightarrow{a.s.} X$, then $E[X_n] \to E[X]$.*

Theorem 4.24. *If X has distribution μ_X on $\mathbb{R}$, then*

$$E[X] = \int_{\mathbb{R}} x\mu_X(dx).$$

Thus to compute $E[X]$ we only need to know the distribution of X. Also, $E[\phi(X)] = \int \phi(x)\mu_X(dx)$ for any Borel measurable function ϕ.

Proposition 4.25 (Key Lemmas and Inequalities)**.** *We have the following key named lemmas and inequalities:*

(0) *(Fubini's Theorem) Let $(\Omega, \mathcal{F}, \mathbb{P}) := (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2)$. Suppose that $F : \Omega \to \mathbb{R}$ is integrable with $F_{\omega_1}(\omega_2) : -F(\omega_1, \omega_2)$. Then*

$$\begin{aligned}\int\int F(w_1, w_2)P_1(dw_1)P_2(dw_2) &= \int\left(\int F_{w_1}(w_2)P_2(dw_2)\right)P_1(dw_1)\\ &= \int\left(\int F_{w_2}(w_1)P_1(dw_1)\right)P_2(dw_1).\end{aligned}$$

(1) *(Markov's Inequality) If $X \geq 0$, then $\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda}E[X]$. Indeed, we see that*

$$\begin{aligned}\mathbb{P}(X \geq \lambda) &= \mathbb{P}(\{X \geq \lambda\}) = E[\chi_{X\geq\lambda}]\\ &= E[\frac{x}{x}\chi_{X\geq\lambda}] \leq E[\frac{x}{\lambda}\chi_{X\geq\lambda}]\\ &= \frac{1}{\lambda}E[x \cdot \chi_{X\geq\lambda}] \leq \frac{1}{\lambda}E[X].\end{aligned}$$

That is, if $X \geq 0$ is integrable, then the tail probability $\mathbb{P}(X \geq \lambda)$ decays at least as $1/\lambda$ for λ large.

(2) *(Chebyshev) If X^2 is integrable and $\lambda > 0$, $\mathbb{P}(|X - E[X]| \geq \lambda) \leq \frac{\mathrm{Var}(X)}{\lambda^2}$. The moral here is that if X^2 is integrable, then the tail probability $\mathbb{P}(X \geq \lambda)$ decays like $1/\lambda^2$ for large $\lambda > 0$.*

(3) *If $X \in L^p$, then $\mathbb{P}(|X| \geq \lambda) \leq \frac{C}{\lambda^p}$. Indeed, since $\forall\lambda > 0$: $\mathbb{P}(|X| \geq \lambda) = \mathbb{P}(|X|^p \geq \lambda^p) \leq \lambda^{-p} \cdot E[|X|^p]$.*

(4) *If $E[e^{\alpha|X|}] < \infty$ for some $\alpha > 0$, then $\mathbb{P}(|X| \geq \lambda) \leq Ce^{-\alpha\lambda}$ for all $\lambda > 0$.*

(5) *(Jensen's Inequality) Let $X = (X_1, \ldots, X_n) : \Omega \to D$ with $D \subset \mathbb{R}^n$ a convex set. If $\phi : D \to \mathbb{R}$ is convex such that $\phi(X)$ is integrable, and if X is integrable, then $E[\phi(X)] \geq \phi(E[X])$.*
Special case: The function *$\phi(x) = |x|^p$ is convex for $p \geq 1$. Then by Jensen's inequality: $E[|X|^p] \geq |E[X]|^p$.*

Proposition 4.26 (Generalized Chebyshev Inequality)**.** *Let $f : [0, \infty) \to [0, \infty)$ be a non-decreasing Borel function, and let $X \geq 0$ be a non-negative random variable. Then for all $a > 0$,*

$$\mathbb{P}\left(X \geq a\right) \leq \frac{E[f(X)]}{f(a)}.$$

Example 4.27 (Spring 2015, #1)**.** Assume X is a symmetric random variable such that $E[X^2] = 1$ and $E[X^4] = 2$. Show that

$$\mathbb{P}(X \geq 1) \leq \frac{14}{27}.$$

Example 4.28 (Spring 2015, #5)**.** for a sequence, $X_1, X_2, \ldots$, of random variables, suppose that we know that

$$\sum_{n\geq 1} nE[|X_n|] < \infty.$$

Show that the sequence $Y_n = X_n + X_{n+1} + \cdots + X_{10n}$ converges almost surely and in L^1 to 0.

Theorem 4.29. *If $X \geq 0$, then*

$$E[X] = \int_0^\infty \mathbb{P}(X \geq \lambda) d\lambda.$$

Corollary 4.30. *If $\mathbb{P}(X > \lambda) \leq \frac{C}{\lambda^{1+\varepsilon}}$ for $\varepsilon > 0$, then X is integrable.*

Proof. We have that

$$\begin{aligned} E[X] &= \int_0^\infty \mathbb{P}(X > \lambda) d\lambda \\ &\leq \int_1^\infty \frac{C}{\lambda^{1+\varepsilon}} d\lambda + \int_0^1 \mathbb{P}(X \geq \lambda) d\lambda \\ &\leq \frac{C}{\varepsilon} + 1 < \infty. \end{aligned}$$

□

Example 4.31. Let $X \sim$ Cauchy so that $f_X(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$. Then

$$\begin{aligned} \mathbb{P}(|X| \geq \lambda) &= \frac{1}{\pi} \int_{-\infty}^{-\lambda} \frac{dx}{1+x^2} + \frac{1}{\pi} \int_\lambda^\infty \frac{dx}{1+x^2} \\ &= \frac{2}{\pi} \int_\lambda^\infty \frac{dx}{1+x^2}, \end{aligned}$$

where

$$\frac{2}{\pi} \int_\lambda^\infty \frac{dx}{2x^2} \leq \frac{2}{\pi} \int_\lambda^\infty \frac{dx}{1+x^2} \leq \frac{2}{\pi} \int_\lambda^\infty \frac{dx}{x^2}.$$

This implies that

$$\frac{2}{2\pi\lambda} \leq \mathbb{P}(|X| > \lambda) \leq \frac{2}{\pi\lambda},$$

which implies that $\lambda \mathbb{P}(|X| > \lambda)$ is constant. But $E[|X|] = +\infty$!

Example 4.32 (Fall 2018, #6)**.** Let $X_1, \ldots, X_n, \ldots$ be identically distributed (but, not necessarily independent) random variables with finite first moment. Is the following,

$$n^{-1} E\left[\max_{1\leq k\leq n} |X_k|\right] \longrightarrow 0,$$

as $n \to \infty$, true or false?

Example 4.33 (Fall 2016, #7)**.** Let X be a random variable taking values on the interval $[1, 2]$. Find sharp lower and upper estimates on the quantity $E[X]E[\frac{1}{X}]$. Provide an example of a random variable for which each of the lower and upper estimates are obtained.

4.5 Variance and covariance

Definition 4.34. We define the *variance* of a random variable X to be

$$\operatorname{Var}(X) := E[(X - E[X])^2] = E[X^2] - E[X]^2, \ \sigma_X := \sqrt{\operatorname{Var}(X)}.$$

For two random variables X, Y we define their covariance to be $\operatorname{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])]$ so that $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$. If $X = (X_1, \ldots, X_n)$ is a random vector, then we define its corresponding *covariance matrix* Cov_X to be the $n \times n$ symmetric matrix whose $(ij)^{th}$ entry is given by $\operatorname{Cov}(X_i, X_j)$. We also define the *correlation* of X and Y to be $\operatorname{Corr}(X, Y) = \operatorname{Cov}(X, Y)/\sigma_X\sigma_Y$.

Proposition 4.35 (Properties)**.** *We have the next properties of the variance and covariance:*

(1) $\operatorname{Var}(X) \geq 0$ *with equality iff* X *is constant almost surely.*

(2) $\operatorname{Var}(X + b) = \operatorname{Var}(X)$, $\forall b \in \mathbb{R}$.

(3) For all $a \in \mathbb{R}$, $\operatorname{Var}(aX) = a^2 \operatorname{Var}(X)$.

(4) $\operatorname{Cov}(X + Z, Y) = \operatorname{Cov}(X, Y) + \operatorname{Cov}(Z, Y)$ *and* $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$.

(5) For any random variables X, Y, $-\sigma_X\sigma_Y \leq \operatorname{Cov}(X, Y) \leq \sigma_X\sigma_Y$, *where equality on the right is attained iff* $Y = \alpha X + \beta$ *for some* $\alpha > 0$*; and on the left iff* $Y = -\alpha X + \beta$ *for some* $\alpha > 0$.

(6) $\operatorname{Corr}(X, Y) = \operatorname{Corr}(aX + b, cY + d)$ *for all* $a, c \neq 0$.

(7) If X, Y *are integrable and independent, then* $E[XY] = E[X]E[Y] \iff \operatorname{Cov}(X, Y) = 0$.

(8) If $X_1, \ldots, X_n$ *are independent and in* L^2*, then* $\operatorname{Var}(a_1X_1 + \cdots + a_nX_n) = a_1^2 \operatorname{Var}(X_1) + \cdots + a_n^2 \operatorname{Var}(X_n)$.

4.6 Uniform integrability

Definition 4.36. The sequence $(X_i)_{i\in I}$ is *uniformly integrable* if

$$\sup_{i\in I} E[|X_i|, |X_i| \geq R] \longrightarrow 0,$$

as $R \to \infty$. The interpretation of this definition is that the tails of the expectation are controlled uniformly.

Proposition 4.37. *If* $(X_i)_{i\in I}$ *are iid and* X_1 *is integrable, then* (X_i) *is uniformly integrable, i.e.,*

$$E[|X_i|, |X_i| \geq R] = E[|X_1|, |X_1| \geq R] \longrightarrow 0,$$

as $R \to \infty$.

Proof. For a fixed $R > 0$, let $Y_R := |X_1|\chi_{|X_1|\geq R}$. Then by dominated convergence, $\forall R > 0$, $|Y_R| \leq |X_1| \implies Y_R \to 0$ as $R \to \infty$. This in turn implies that $E[Y_R] \to 0$ as $R \to \infty$. □

Proposition 4.38. *If* $\exists\, Y$ *such that* $|X_i| \leq Y$ *for all* $i \geq 1$ *with* Y *integrable, then* (X_i) *is uniformly integrable.*

Proof. Since $\{|X_i| \geq R\} \subseteq \{Y \geq R\}$, we have the argument that

$$E[|X_i|, |X_i| \geq R] \leq E[Y, |X_i| \geq R] \leq E[Y, Y \geq R] \longrightarrow 0,$$

as $R \to \infty$ (by integrability). So $\sup_{i\in I} E[|X_i|, |X_i| \geq R] \to 0$ as $R \to \infty$, as claimed. □

Proposition 4.39. *If* $\sup_{i \in I} E[|X_i|^\alpha] < \infty$ *for some* $\alpha > 1$*, then* $(X_i)_{i \in I}$ *is uniformly integrable.*

Proof. We see that

$$E[|X_i|, |X_i| \geq R] = E[|X_i| \chi_{|X_i| \geq R}] \leq E\left[|X_i| \left(\frac{|X_i|}{R}\right)^{\alpha-1} \chi_{|X_i| \geq R}\right]$$
$$= E\left[\frac{|X_i|^\alpha}{R^{\alpha-1}} \chi_{|X_i| \geq R}\right] \leq \frac{E[|X_i|^\alpha]}{R^{\alpha-1}},$$

which implies that

$$\sup_{i \in I} E[|X_i|, |X_i| \geq R] \leq \sup_{i \in I} \frac{E[|X_i|^\alpha]}{R^{\alpha-1}} \longrightarrow 0,$$

as $R \to \infty$. $\square$

4.7 Convergence and expectation

First, some preliminary notes: we recall the subset containment properties of limsups of sets given in Proposition 1.4. Also, we restate the precise definition of an *exponential random variable* here for clarity of exposition as: $X \sim \mathrm{Exp}(\lambda)$ is exponential is μ_X has a density function of the form

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Now we arrive at the focus of this subsection which is to dig into the next problem on iid exponential random variables:
Problem Statement: We claim that if $X_1, \ldots, X_n, \ldots$ are iid with $X_1 \sim \mathrm{Exp}(1)$, then

$$\mathbb{P}\left(\limsup_{n \to \infty} \frac{X_n}{\log n} = 1\right) = 1$$

We notice that a proof of this result is equivalent to showing that

(🐧) $\mathbb{P}\left[\limsup_{n\to\infty} \frac{X_n}{\log n} > 1 + \delta\right] = 0$, $\forall \delta > 0$; and then similarly that

(☕) $\mathbb{P}\left[\limsup_{n\to\infty} \frac{X_n}{\log n} > 1 - \delta\right] = 1$, $\forall \delta > 0$.

Note next that

$$\left\{\limsup_{n \to \infty} \frac{X_n}{\log n} = 1\right\} = \bigcap_{\delta > 0} \left\{1 + \delta > \limsup_{n \to \infty} \frac{X_n}{\log n} > 1 - \delta\right\} = \bigcap_{k \geq 1} A_k,$$

where

$$A_k := \left\{1 + \frac{1}{k} > \limsup_{n \to \infty} \frac{X_n}{\log n} > 1 - \frac{1}{k}\right\},$$

for $k \geq 1$. Also, we have that $A_k \searrow \left\{\limsup_{n\to\infty} \frac{X_n}{\log n} = 1\right\}$. Moving along, we can see that

$$\mathbb{P}\left[\frac{X_n}{\log n} > 1 + \delta - \varepsilon\right] = e^{-(1+\delta-\varepsilon)\log n} = \frac{1}{n^{1+\delta-\varepsilon}}$$
$$\Longrightarrow \sum_{n \geq 1} \mathbb{P}\left[\frac{X_n}{\log n} > 1 + \delta - \varepsilon\right] = \sum_{n \geq 1} \frac{1}{n^{1+\delta-\varepsilon}} < \infty, \quad \text{if } \varepsilon < \delta.$$

So in summary, to complete the proof of 🐧, we use Borel-Cantelli in the form of

$$\mathbb{P}\left[\left\{\limsup_{n\to\infty}\frac{X_n}{\log n}>1+\delta\right\}\right]\subseteq\mathbb{P}\left[\limsup_{n\to\infty}\left\{\frac{X_n}{\log n}>1+\delta\right\}\right]=0.$$

To show ☕, we need to consider lower bounds: $\forall\varepsilon>0$,

$$\limsup_{n\to\infty}\left\{\frac{X_n}{\log n}>1-\delta+\varepsilon\right\}\subseteq\left\{\limsup_{n\to\infty}\frac{X_n}{\log n}>1+\delta\right\}.$$

To use the second part of Borel-Cantelli on independent random variables, we want that

$$\sum_{n\geq 1}\mathbb{P}\left[\frac{X_n}{n}>1-\delta+\varepsilon\right]=\sum_{n\geq 1}\frac{1}{n^{1-\delta+\varepsilon}}=+\infty,$$

which happens if $\varepsilon\leq\delta$. Then this implies that

$$\mathbb{P}\left[\limsup_{n\to\infty}\left\{\frac{X_n}{\log n}>1+\delta-\varepsilon\right\}\right]=1,\quad\text{for }\varepsilon\leq\delta\implies\text{☕}.$$

Example 4.40 (Spring 2018, #1)**.** Let $(X_n)_{n\geq 1}$ be a sequence of iid random variables with $X_1\sim\mathrm{Exp}(\lambda)$, i.e., for $t\geq 0$ we have that $\mathbb{P}(X_n\geq t)=e^{-\lambda t}$ where $\lambda>0$. Prove that

$$\limsup_{n\to\infty}\frac{X_n}{\log n}<\infty,\ \text{a.s.}$$

4.8 Characteristic functions of a random variable

Definition 4.41 (Characteristic Function)**.** The *characteristic function* of a random varible X is the function $\phi_X:\mathbb{R}\to\mathbb{C}$ defined by $\phi_X(t):=E[e^{\imath Xt}]$.

We have the following properties of the characteristic function ϕ_X of any random variable X:

- We have that $\phi_X(0)=1$ and for all t: $|\phi_X(t)|=|E[e^{\imath Xt}]|\leq E[|e^{\imath Xt}|]=1$;
- The characteristic function $\phi_X(t)$ is a uniformly continuous function of t.
- If $Y=aX+b$ for constants $a,b\in\mathbb{R}$, then
$$\phi_Y(t)=E[e^{\imath(aX+b)t}]=e^{\imath bt}\phi_X(at).$$
- The function $M_X(t)=e^{tX}$ is called the *moment generating function* of X. Unlike the characteristic function, this GF does not necessarily exist for all $t\in\mathbb{R}$. For those t such that $M_X(t)$ does converge, we have the relation that $\phi_X(t)=M_X(\imath t)$;
- If X has a density f, then
$$\phi_X(t)=\int_{-\infty}^{\infty}e^{\imath tx}f(x)dx,$$
which is the *Fourier transform of* f at $-t$;
- If two random variables have the same characteristic function, then they have the same underlying distribution;
- If $X_1,\dots,X_n$ are independent random variables, then
$$\phi_{X_1+\cdots+X_n}(t)=\phi_{X_1}(t)\cdots\phi_{X_n}(t).$$

- If $X \sim N(0,1)$, then $\phi_X(t) = e^{-t^2/2}$. Similarly, if $Y \sim N(\mu, \sigma^2)$, then Y has the same distribution as $\sigma X + \mu$, and hence, $\phi_Y(t) = e^{\imath \mu t} e^{-(\sigma t)^2/2}$.
- If the characteristic functions of X_n converge pointwise to that of X, then $X_n \implies X$ weakly.
- If X, Y are independent, then the characteristic function of $X+Y$ equals the product of their characteristic functions: $\phi(u) = \phi_X(u)\phi_Y(u)$.

Proposition 4.42 (Moments of a Random Variable)**.** *Let ϕ be the characteristic function of X. If $E[|X|] < \infty$, then ϕ is continuously differentiable, and*

$$\phi'(0) = \imath E[X].$$

If $E[|X|^k] < \infty$ for some positive integer $k \geq 1$, then ϕ has k continuous derivatives and

$$\phi^{(j)}(0) = \imath^j E[X^j], \ \forall 0 \leq j \leq k.$$

Example 4.43 (Spring 2018, #9)**.** Find an example of a random variable with a density function, but whose characteristic function ϕ_X satisfies

$$\int_{-\infty}^{\infty} |\phi_X(t)| dt = \infty.$$

Example 4.44 (Fall 2018, #4)**.** Let ϕ be the characteristic function of a random variable X. Show that

$$\Psi_1(t) = |\phi(t)|^2, \Psi_2(t) = \frac{1}{t}\int_0^t \phi(s) ds,$$

are also characteristic functions.

Example 4.45 (Spring 2017, #1)**.** Show that if X_n and Y_n are independent for $n = 1, 2, \ldots$ and $X_n \implies X$ and $Y_n \implies Y$, where X, Y are independent, then $X_n + Y_n \implies X + Y$.

Example 4.46 (Spring 2017, #5)**.** Show that the characteristic function ϕ of a random variable is real if and only if X and $-X$ have the same distribution.

Example 4.47 (Fall 2016, #5)**.** Let ξ_1, ξ_2 be independent random variables with respective characteristic functions

$$\phi_1(u) = \frac{1 - \imath u}{1 + u^2}, \phi_2(u) = \frac{1 + \imath u}{1 + u^2}.$$

Find the probability that $\xi_1 + \xi_2$ takes values in $(3, \infty)$.

Example 4.48 (Spring 2016, #5)**.** Let $(N_t)_{t \geq 0}$ be a rate-λ Poisson process. Let $X_1, X_2, \ldots$ be iid random variables with $E[X_1] < \infty$, and define

$$S_t := \sum_{i=1}^{N_t}.$$

Show that S_t/t converges in probability to a constant and compute this constant.

5 Modes of convergence

5.1 Definitions of convergence

Definition 5.1 (Convergence Everywhere)**.** Let $X_n, X : \Omega \to \mathbb{R}$. Then $X_n \to X$ if $X_n(\omega) \to X(\omega)$ as $n \to \infty$ for all $\omega \in \Omega$.

Definition 5.2 (Almost Sure Convergence)**.** We say that an event $A \subseteq \Omega$ is *almost sure* if $\mathbb{P}(A) = 1$, while noting that this property is distinguished from "provably certain" in so much as there may be other outcomes in the sample space which lie outside of A, but these events are in zero measure subsets of the field. Now, we write $X_n \xrightarrow{a.s.} X$ if $\mathbb{P}(\{\omega : X_n(\omega) \to X(\omega)\}) = 1$. Equivalently, we define that $X_n \xrightarrow{a.s.} X$ if $\exists A \in \mathcal{F}$ an almost sure event such that $X_n(\omega) \to X(\omega)$, $\forall \omega \in A$. We have almost sure convergence if

$$\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n,$$

in which case we may write that

$$X = \lim_{n\to\infty} X_n = \limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n.$$

Proposition 5.3. $X_n \xrightarrow{a.s.} X \iff \forall \varepsilon > 0$, $\mathbb{P}(\limsup_{n\to\infty}\{|X_n - X| > \varepsilon\}) = 0$. *An important corollary is that if* $\forall \varepsilon > 0$ $\sum_{n\geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, *then* $X_n \xrightarrow{a.s.} X$.

Definition 5.4 (Convergence in Probability)**.** We write $X_n \xrightarrow{\mathbb{P}} X$ if $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$. Note that convergence in probability implies finiteness (a.e.) almost surely.

Definition 5.5 (Convergence in Mean, a.k.a., L^p Convergence)**.** We write $X_n \xrightarrow{L^p} X$ if $E[|X_n - X|^p] \to 0$ as $n \to \infty$.

Definition 5.6 (Weak Convergence, a.k.a., Convergence in Distribution)**.** A sequence of random variables $(X_n)_{n\geq 1}$ *converges in distribution*, or *converges weakly*, to X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

for every $x \in \mathbb{R}$ at which F_X is continuous. In this case, we write $X_n \xrightarrow{D} X$, or sometimes also $X_n \xrightarrow{\text{dist}} X$.

Proposition 5.7 (Equivalent Conditions for Convergence in Distribution)**.** *For any sequence* $(X_n)_{n\geq 1}$ *of random variables, we have that* $X_n \xrightarrow{\text{dist}} X$ *if and only iff any of the following conditions are met:*

(a) $E[f(X_n)] \longrightarrow E[f(X)]$, *for all bounded, continuous functions* f*;*

(b) $E[f(X_n)] \longrightarrow E[f(X)]$, *for all bounded, Lipschitz functions* f*. Here, we say that a function* $f : \mathbb{R} \to \mathbb{R}$ *is Lipschitz, or 1–Lipschitz, if there exists a constant* $K > 0$ *which is independent of* $x, y \in \mathbb{R}$ *such that* $\forall x, y$: $|f(x) - f(y)| \leq K \cdot |x - y|$*;*

(c) $\limsup_{n\to\infty}\{\mathbb{P}(X_n \in C)\} \leq \mathbb{P}(X \in C)$, *for all closed sets* $C \subseteq \Omega$.

Example 5.8 (Fall 2018, #7)**.** Let $X_1, X_2, \ldots$ be iid random variables with common characteristic function ϕ, and let $S_n = \sum_{k=1}^{n} X_k$. Show that if ϕ is differentiable at 0 with $\phi'(0) = \imath\mu$, then, as $n \to \infty$, $S_n/n \to \mu$ in probability.

Example 5.9 (Fall 2016, #8)**.** Show that for a sequence of random variables X_n, one has that $X_n \xrightarrow{\mathbb{P}} X$ in probability if and only if

$$E\left[e^{\min(2,|X_n-X|)}-1\right] \longrightarrow 0,$$

as $n \to \infty$.

Example 5.10 (Spring 2016, #4)**.** Let $X_1, X_2, \ldots$ be iid exponential variables with parameter $\lambda := 1$, and set

$$M_n := \max(X_1, \ldots, X_n).$$

Find sequences $(a_n), (b_n) \subset \mathbb{R}$ such that $(M_n - a_n)/b_n$ converges in distribution.

Example 5.11 (Spring 2015, #6)**.** Assume $(X_n)_{n\geq 1}$ is a sequence of iid random variables with mean 0 and variance 1. Show that

$$Y_n = \frac{\sqrt{n}X_1 + \sqrt{n-1}X_2 + \cdots + X_n}{n},$$

converges weakly (in distribution) to a normal $N(0, 1/2)$.

Example 5.12 (Spring 2015, #7)**.** Assume that $(U_n)_{n\geq 1}$ is a sequence of uniform random variables on $[0,1]$. Let $V_n := \max(U_1, U_2^2, \ldots, U_n^n)$. Show that $(1 - V_n)\log(n)$ converges weakly (in distribution) to an exponential random variable with parameter $\lambda = 1$.

5.2 Remarks on the equivalence of studying convergence to zero

Proposition 5.13 (Convergence to Zero)**.** *If $(Y_n)_{n\geq 1}$ is any sequence of random variables, then $Y_n \xrightarrow{a.s.} 0 \iff \forall\varepsilon > 0,\ \mathbb{P}(\limsup |Y_n| > \varepsilon) = 0$. And then as a consequence, $X_n \xrightarrow{a.s.} X \iff \forall\varepsilon > 0,\ \mathbb{P}(\limsup |X_n - X| > \varepsilon) = 0$.*

Remark 5.14 (Notes on Convergence to Zero)**.** We notice that it is sufficient to study convergence of a random variable to zero since we can always replace the expression $X_n \xrightarrow{a.s.} X$ by $Y_n := X_n - X \xrightarrow{a.s.} 0$. Next, we have the observation that for any sequence $(a_n)_{n\geq 1}$,

$$\begin{aligned} a_n \longrightarrow 0 &\iff \limsup_{n\to\infty} |a_n| = 0 \\ &\iff \inf_{n\geq 1} \sup_{k\geq n} a_k = 0 \\ &\iff \forall\varepsilon > 0, \exists N_\varepsilon : \sup_{k\geq N_\varepsilon} a_k < \varepsilon. \end{aligned}$$

Remark 5.15 (Borel-Cantelli as a Practical Tool for Proving Convergence)**.** As a tool for proving almost sure convergence, we can apply Borel-Cantelli: If $\forall\varepsilon > 0$, $\sum_{n\geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, then $X_n \xrightarrow{a.s.} X$. Note that the convergence of the above sum, at minimum requires that $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$ for all $\varepsilon > 0$. We also point out the following key observations:

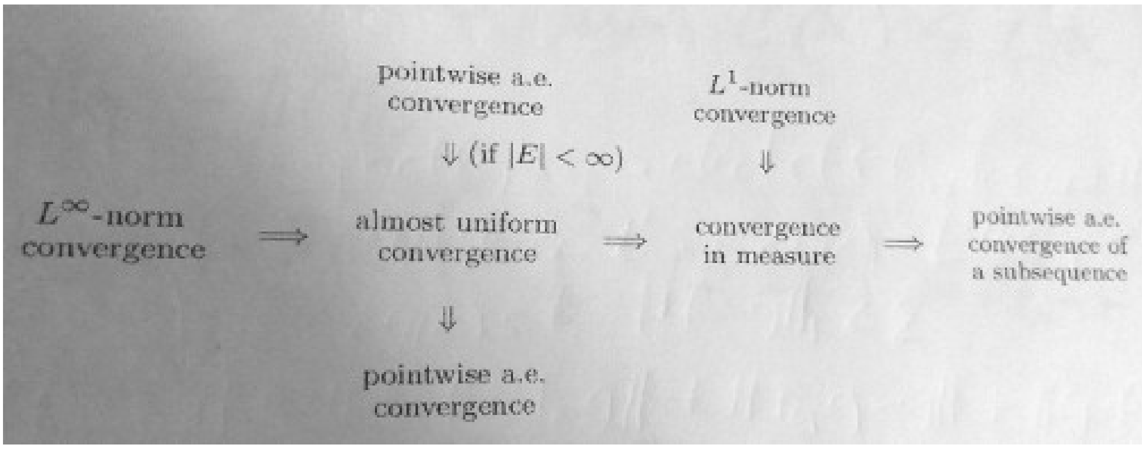
- If $\forall\varepsilon > 0$: $\sum_{n\geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, then $\mathbb{P}(\limsup\{|X_n - X| > \varepsilon\}) = 0$, for all $\varepsilon > 0$;
- If $\sum_{n\geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, $\forall\varepsilon > 0$, then $X_n \xrightarrow{a.s.} X$, where we consider $\mathbb{P}(|X_n - X| > \varepsilon)$ to be the *tail probability* we must estimate for n large.

5.3 Flowcharts of implications: convergence hieracrhchy

We have the following map of convergence implications:

$$\begin{array}{ccccc} \xrightarrow{L^s} & \underset{s>r\geq 1}{\Rightarrow} & \xrightarrow{L^r} & & \\ & & \Downarrow & & \\ \xrightarrow{\text{a.s.}} & \Rightarrow & \xrightarrow{\mathbb{P}} & \Rightarrow & \xrightarrow{\text{Dist}} \end{array} \quad \text{(Hierarchy of the Modes of Convergence)}$$

We compare this to the real analysis modes summary chart reproduced as follows from Heil's notes:



Proposition 5.16 (Almost Everywhere Convergence of a Subsequence)**.** *Convergence in probability implies that there exists a subsequence (k_n) of the original sequence which almost surely converges:*

$$\exists \{k_n\}_{n\geq 1} \subseteq \mathbb{N} : X_n \xrightarrow{\mathbb{P}} X \implies X_{k_n} \xrightarrow{a.s.} X.$$

Disproof: Counterexample to Convergence in Probability Implies Almost Sure Convergence. We will demonstrate a sequence of random variables that converges in probability to zero on $[0,1]$, but which does not converge almost surely to zero. For $n \geq 1$ and $1 \leq j \leq n$, define

$$S_{n,j} := \left[\frac{j-1}{n}, \frac{j}{n}\right] \subseteq [0,1],$$

and set $X_n(x) := \chi_{S_{n,j}}(x)$. Let $\varepsilon > 0$ and observe that

$$\mathbb{P}\left(\{x \in [0,1] : f_n(x) > \varepsilon\}\right) = \mathbb{P}\left(\{x \in [0,1] : f_n(x) = 1\}\right) = |S_{n,j}| = \frac{1}{n} \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

On the other hand, given any $x \in [0,1]$ there are infinitely-many $n \in \mathbb{N}$ such that $x \in S_{n,j}$, i.e., such that $X_n(x) = 1$. This implies that there is a subsequence $\{X_{n_k}\}$ such that $X_{n_k} \xrightarrow{\mathbb{P}} 1$ and hence $X_n \nrightarrow 0$ almost surely on $[0,1]$. □

Proof: Existence of an Almost Surely Convergent Subsequence. Suppose that $X_n \xrightarrow{\mathbb{P}} X$. We need to find a subsequence $\{X_{n_k}\}_{k\geq 1}$ such that $X_{n_k} \xrightarrow{a.s.} X$ as $k \to \infty$ for all $\omega \in A$ where $A \subseteq \Omega$ is some large enough event in the sample space such that $\mathbb{P}(A) = 1$. For $j \geq 1$, choose L_j such that for all $k \geq L_j$

$$\mathbb{P}\left(\{\omega : |X_k - X|(x) \geq 1/j\}\right) < \frac{1}{2^j}.$$

We can just as well assume here that the L_j's satisfy: $L_1 < L_2 < L_3 < \cdots$. For $j \geq 1$, we define

$$E_j := \{\omega : |X_{L_j} - X|(\omega) \geq 1/j\},$$

and set

$$Z := \limsup_{j\to\infty} E_j = \bigcap_{m\geq 1}\bigcup_{j\geq m} E_j.$$

Now for $m \geq 1$, we can see that

$$\mathbb{P}(Z) \leq \mathbb{P}\left(\bigcup_{j\geq m} E_j\right) \leq \sum_{j\geq m}\mathbb{P}(E_j) < \sum_{j\geq m} 2^{-j} = 2^{1-m} \longrightarrow 0,$$

as $m \to \infty$. So we conclude that $\mathbb{P}(Z) = 0$. Next, if $\omega \in \Omega \setminus Z$, then $\omega \notin \cup_{j\geq m} E_j$ for some $m \geq 1$. This implies that $x \notin E_j$ for all $j \geq m$. Thus $|X_{L_j} - X|(\omega) \leq 1/j$ for $j \geq m$, which implies that $\lim_{j\to\infty} X_{L_j}(\omega) = X(\omega)$ for all $x \in \Omega \setminus Z$. So it suffices to take $X_{n_j} := X_{L_j}$ so that $X_{n_k} \xrightarrow{a.s.} X$, or equivalently with $A := \Omega \setminus Z$ so that $X_{n_k}(\omega) \longrightarrow X(\omega)$ for all $\omega \in A$ where, as proved above, $\mathbb{P}(A) = \mathbb{P}(\Omega) - \mathbb{P}(Z) = 1 - 0 = 1$. $\square$

5.4 Convergence theorems (analogs to real analysis)

Theorem 5.17 (Lebesgue's BCT)**.** *Let f_n be integrable $\forall n \geq 1$ such that (a) $f_n \xrightarrow{\mathbb{P}} f$; or (b) $f_n \xrightarrow{a.s.} f$ for some measurable function f. If $|f_n(x)| \leq g(x)$ a.e. for all n where g is integrable, then f is integrable and*

$$\lim_{n\to\infty}\int |f_n - f|\,d\mu = 0.$$

Theorem 5.18. *If f, g are measurable such that $|f| \leq g$ and g is integrable, then f is integrable.*

Theorem 5.19 (Monotone convergence theorem)**.** *Let $\{f_k\}_{k\geq 1}$ be a sequence of measurable functions on E. Then*

(1) If $f_k \nearrow f$ a.s. on E and $\exists\varphi$ integrable such that $f_k \geq \varphi$ a.e. in E for all $k \geq 1$, then

$$\lim_{k\to\infty}\int_E f_k = \int_E f.$$

(2) If $f_k \searrow f$ a.s. on E and $\exists\varphi$ integrable such that $f_k \leq \varphi$ $\forall k \geq 1$, then

$$\lim_{k\to\infty}\int_E f_k = \int_E f.$$

Theorem 5.20 (General form of LDCT)**.** *Let f_k be measurable for all $k \geq 1$. Suppose that (a) $\lim_{k\to\infty} f_k = f$ a.e. in E; and (b) $\exists\varphi$ integrable such that for all $k \geq 1$, $|f_k| \leq \varphi$ a.e. in E. Then*

$$\lim_{k\to\infty}\int f_k = \int f.$$

Lemma 5.21 (Fatou)**.** *Suppose that $E \subseteq \mathbb{R}^n$ is measurable and let $f_k \geq 0$ be measurable for all $k \geq 1$. Then*

$$\liminf_{k\to\infty}\int_E f_k \geq \int_E \left(\liminf_{k\to\infty} f_k\right).$$

We can obtain the same conclusion if we instead assume that the $f_k \geq \varphi$ for all $k \geq 1$ when φ is integrable on E. Notice that the statement of Fatou's lemma makes no a priori assumptions on the convergence of the sequence $\{f_k\}_{k\geq 1}$.

Theorem 5.22 (Uniform convergence theorems)**.** *We have the following variants of uniform convergence theorems:*

- $f_n \to f$ *uniformly on* $[a,b]$ *with* f_n *all continuous* $\implies$ f *is continuous.*
- *If* f_n *is differentiable on* $[a,b]$ *and* $\lim_{n\to\infty} f_n(x_0)$ *exists for some* $x_0 \in [a,b]$ *and* f_n' *converge uniformly on* $[a,b]$*, then* $f_n \to f$ *uniformly on* $[a,b]$ *and* $f'(x) = \lim_{n\to\infty} f_n'(x)$ *for all* $x \in [a,b]$.
- *If* f_n *is integrable on* $[a,b]$ *and* $f_n \to f$ *uniformly, then* f *is integrable and*

$$\int_{[a,b]} f = \lim_{n\to\infty} \int_{[a,b]} f_n.$$

5.5 Properties of the modes of convergence for random variables

Fact: If $\mu(X) < \infty$, then $L^q(\mu) \subsetneq L^p(\mu)$ for any $0 < p < q \leq \infty$. This can be proved by applying Hölder with the conjugate exponents $\boxed{p_0 := q/p \text{ and } q_0 := q/(q-p)}$.

Proposition 5.23 (Cauchy-Schwarz and Hölder inequalities)**.** *Note that we assume that the functions* f, g *are both square integrable, i.e.,* $|f|^2, |g|^2$ *are both integrable. If this is the case then,*

$$\int_0^1 |f| \leq \left(\int_0^1 |f|^2\right)^{1/2}$$
$$\left|\int fg\right| \leq \left(\int |f|^2\right)^{1/2} \left(\int |g|^2\right)^{1/2}.$$

Note that the latter equation above is the special case of the more general cases in Hölder's inequality for $\frac{1}{p} + \frac{1}{q} = 1 \iff p + q = pq$ *when* $p = q = 2$*:*

$$\left|\int fg\right| \leq ||f||_p \cdot ||g||_q.$$

Dual (conjugate) exponents: $\boxed{\textit{If } p+q = pq, \textit{ then } p = q/(q-1).}$

Note that *Minkowski's theorem* states that $||f+g||_p \leq ||f||_p + ||g||_p$. Minkowski also implies that

$$\left\|\int f\right\|_p \leq \int ||f||_p.$$

Proof of Minkowski. We let p be arbitrary with $f, g \in L^p$ and observe that:

$$\begin{aligned}
||f+g||_p^p &= \int |f+g|^p = \int |f+g|^{p-1}|f+g| \\
&\leq \int |f+g|^{p-1}(|f|+|g|)d\mu, \quad \text{then we apply Hölder} \\
&\leq (||f||_p + ||g||_p)\left(\int |f+g|^{(p-1)\frac{p}{p-1}}\right)^{1-1/p} \\
&= (||f||_p + ||g||_p)\frac{||f+g||_p^p}{||f+g||_p}.
\end{aligned}$$

□

If $\alpha > 0$, we can obtain that

$$\mu\left(\{x : |f(x)| > \alpha\}\right) \leq \left(\frac{||f||_p}{\alpha}\right)^p.$$

Another important result which is non-trivial to prove and due to Riesz is that for $1 \leq q < \infty$ and $g \in L^1$:

$$||g||_q = \sup\left\{\left|\int fg\right| : ||f||_p = 1\right\}, \quad \text{where } \frac{1}{p} + \frac{1}{q} = 1.$$

Example 5.24 (Dense Class Arguments)**.** *Compactly supported continous functions are dense in* L^1*:* $C_c[a,b] = L^1[a,b]$. Other dense functions in L^p are the simple (i.e., staircase) functions, and polynomials. If $f \in L^p$ and C is *dense* in L^p then $\forall \varepsilon > 0$, $\exists g \in C$ such that $||f - g||_p < \varepsilon$.

Proposition 5.25. *Provided that the probability space is complete:*

- *If* $X_n \xrightarrow{P} X$ *and* $X_n \xrightarrow{P} Y$*, then* $X = Y$ *almost surely.*
- *If* $X_n \xrightarrow{a.s.} X$ *and* $X_n \xrightarrow{a.s.} Y$*, then* $X = Y$ *almost surely.*
- *If* $X_n \xrightarrow{L^r} X$ *and* $X_n \xrightarrow{L^r} Y$*, then* $X = Y$ *almost surely.*
- *If* $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$*, then for any* $a, b \in \mathbb{R}$, $aX_n + bY_n \xrightarrow{P} aX + bY$*, and* $X_nY_n \xrightarrow{P} XY$*.*
- *If* $X_n \xrightarrow{a.s.} X$, $Y_n \xrightarrow{a.s.} Y$*, then for any* $a, b \in \mathbb{R}$, $aX_n + bY_n \xrightarrow{a.s.} aX + bY$*, and* $X_nY_n \xrightarrow{a.s.} XY$*.*
- *If* $X_n \xrightarrow{L^r} X$, $Y_n \xrightarrow{L^r} Y$*, then for any* $a, b \in \mathbb{R}$, $aX_n + bY_n \xrightarrow{L^r} aX + bY$*.*
- *None of the above statements are true for weak convergence in distribution.*
- *Convergence in distribution to a constant implies convergence in probability.*

6 Laws of large numbers

6.1 The weak law of large numbers

Example 6.1 (Key Observations)**.** Suppose that $(X_i)_{i\geq 1}$ have the same L^2 norm and that $E[X_i]=0$ for all $i\geq 1$. Then:

- $\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = E\left[\left|\frac{X_1+\cdots+X_n}{n}\right|^2\right]$, which implies that

$$\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = \left\|\frac{X_1+\cdots+X_n}{n}\right\|_{L^2}^2 \leq \frac{1}{n}\left(||X_1||^2+\cdots+||X_n||^2\right) = ||X_1||_{L^2}^2.$$

- If $X_1=\cdots=X_n$, then $\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = \mathrm{Var}(X_1)$.
- If $X_1,\ldots,X_n$ are iid, then

$$\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{1}{n^2}\sum_{i,j=1}^{n}\mathrm{Cov}(X_i,X_j) = \frac{1}{n^2}\sum_{m=1}^{n}\mathrm{Var}(X_m) = \frac{1}{n}\mathrm{Var}(X_1).$$

Theorem 6.2 (WLLN)**.** *If $X_1,\ldots,X_n$ are iid with mean μ and variance σ^2, then $\bar{S}_n := \frac{X_1+\cdots+X_n}{n} \xrightarrow{P,L^2} \mu$ as $n\to\infty$.*

Proof. First, we show the statement of L^2 convergence. Since $\mu = E[\bar{S}_n]$:

$$E[(\bar{S}_n-\mu)^2] = \mathrm{Var}(\bar{S}_n) = \frac{\sigma^2}{n} \to 0,$$

as $n\to\infty$. So $\bar{S}_n \xrightarrow{L^2} \mu$. Next, for all $\varepsilon>0$,

$$\mathbb{P}(|\bar{S}_n-\mu|>\varepsilon) \leq \frac{E[(\bar{S}_n-\mu)^2]}{\varepsilon^2} \leq \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0,$$

as $n\to\infty$. Hence, $\bar{S}_n \xrightarrow{P} \mu$. □

Theorem 6.3 (WLLN, V2)**.** *Suppose that $X_1,\ldots,X_n$ are iid with mean μ and $E[X_1^4]<\infty$. Then $\bar{S}_n \xrightarrow{L^4,a.s.} \mu$.*

Proof. Since we can replace X_i by $Y_i := X_i-\mu$ and bound the resulting norm of this random variable, we can assume that $\mu=0$. Using Markov,

$$\mathbb{P}(|\bar{S}_n|>\varepsilon) = \mathbb{P}(\bar{S}_n^4>\varepsilon^4) \leq \frac{E[\bar{S}_n^4]}{\varepsilon^4} \leq \frac{C}{n^2\varepsilon^4}.$$

Thus $\bar{S}_n \xrightarrow{P} 0$. And by Borel-Cantelli, we have that

$$\sum_{n\geq 1}\mathbb{P}(|\bar{S}_n|\geq\varepsilon) \leq \sum_{n\geq 1}\frac{C}{n^2\varepsilon^4} < \infty,$$

and so $\bar{S}_n \xrightarrow{a.s.} 0$. □

Theorem 6.4. *If $X_1,\ldots,X_n$ are independent with mean μ and are uniformly integrable, then $\bar{S}_n \xrightarrow{L^1,P} \mu$.*

6.2 Large deviations

If $X_1, \ldots, X_n$ are iid with mean 0 and $E[e^{\alpha|X_1|}] < \infty$ for some $\alpha > 0$, then

$$\mathbb{P}(|\bar{S}_n| > \varepsilon) \leq 2e^{-nI(\varepsilon)},$$

where the *rate function* $I(\varepsilon)$ is given by

$$I(\varepsilon) = \sup_{\lambda \in \mathbb{R}} \left\{ \varepsilon\lambda - \log E[e^{\lambda X_1}] \right\}.$$

The function $f(\lambda) := \log E[e^{\lambda X_1}]$ is a convex function with

$$f(0) = 0, f'(0) = E[X_1], f''(0) = \frac{E[X_1^2 e^{\lambda X_1}]E[e^{\lambda X_1}] - E[X_1 e^{\lambda X_1}]^2}{E[e^{\lambda X_1}]^2} \geq 0.$$

Also, the rate function is convex and satisfies

$$I(0) = 0, I'(0) = 0, I''(0) = \frac{\sigma^2}{2} = \frac{\mathrm{Var}(X_1)}{2}.$$

We remark that if $I(a) < \infty$, then $\bar{S}_n$ "visits" $(a - \varepsilon, a + \varepsilon)$ with positive probability.

Theorem 6.5. *The moment generating function of a random variable X is defined to be $M(\lambda) := E[e^{\lambda X}]$. Define*

$$I(a) := \sup_{\lambda \in \mathbb{R}} \{a\lambda - \log M(\lambda)\}.$$

Then

- *(1) $\mathbb{P}(\bar{S}_n \geq a) \leq e^{-nI(a)}$, for $a > \mu$;*
- *(2) $\mathbb{P}(\bar{S}_n \leq a) \leq e^{-nI(a)}$, for $a < \mu$;*
- *(3) We have that*

$$\mathbb{P}(|\bar{S}_n - a| < \varepsilon) \geq \left(\frac{1 - F(a)}{n\varepsilon^2} \right) e^{-n(I(a) + \varepsilon G(a))},$$

where the functions F, G are explicit in terms of μ, I, M.

Theorem 6.6 (WLLN, V3)**.** *If $(X_i)_{i \geq 1}$ are uncorrelated with mean μ such that $\sigma^2 = \sup_{i \geq 1} \mathrm{Var}(X_i) < \infty$, then $\bar{S}_n \xrightarrow{P, L^2} \mu$ as $n \to \infty$.*

Theorem 6.7 (WLLN, First Improvement, V4)**.** *If $(X_i)_{i \geq 1}$ are iid, $\mu = E[X_i]$, and $E[X_1^4] < \infty$, then $\bar{S}_n \xrightarrow{P, L^4, a.s.} \mu$ as $n \to \infty$.*

Proof. This proof is instructive. We assume that we have $\mu = 0$, as otherwise we can take $\widetilde{X}_i := X_i - \mu$ and notice that $||\widetilde{X}_i||_{L^4} \leq \mu + ||X_i||_{L^4} < \infty$. Next, we can expand

$$E[\bar{S}_n^4] = \frac{1}{n^4} E[(X_1 + \cdots + X_n)^4] = \frac{1}{n^4} \sum_{i,j,k,\ell=1}^{n} E[X_i X_j X_k X_\ell].$$

Observe that if one of the indices i, j, k, ℓ appears by itself in the previous expansion, then $E[X_i X_j X_k X_\ell] = 0$ (by independence). The surviving terms are of the forms $(ij\hat{k}\hat{\ell}), (\hat{i}j\hat{k}\ell), (\hat{i}jk\hat{\ell})$ and $i = j = k = \ell$. In the first three cases, we obtain terms of $E[X_1^2]^2$, and in the last case we obtain $E[X_1^4]$. This implies that

$$E[\bar{S}_n^4] = \frac{1}{n^4} \left(3n(n-1)E[X_1^2]^2] + nE[x_1^4] \right).$$

Now since $\mathrm{Var}(X_1^2) = E[X_1^4] - E[X_1^2]^2 \geq 0$, we have that

$$E[\bar{S}_n^4] \leq \frac{1}{n^4}(3n(n-1)+n)E[X_1^4] \leq \frac{3E[X_1^4]}{n^2}.$$

Then this implies that

$$\mathbb{P}(|\bar{S}_n| \geq \varepsilon) = \mathbb{P}(\bar{S}_n^4 \geq \varepsilon^4) \leq \frac{3E[X_1^4]}{n^2\varepsilon^4} \longrightarrow 0,$$

as $n \to \infty$. And also, by Borel-Cantelli, we have that $\bar{S}_n \xrightarrow{a.s.} 0$ as $n \to \infty$. □

Remark 6.8 (Observations and Extensions)**.** If (X_i) are independent, $E[X_i] = 0$ for all $i \geq 1$, and $\sup_{i\geq 1} E[X_i^4] < \infty$, then $\bar{S}_n \xrightarrow{P,L^4,a.s.} 0$. If $(X_i)_{i\geq 1}$ are independent, $E[X_i] = 0$ $\forall i$, and $\sup_{i\geq 1} E[X_i^{2k}] < \infty$, then $\bar{S}_n \xrightarrow{P,L^{2k},a.s.} 0$. In fact, we can show that $\mathbb{P}(|\bar{S}_n| \geq \varepsilon) \leq \frac{C_k}{(n\varepsilon^2)^k}$, where the constant C_k depends only on the fixed $k \geq 1$.

6.3 The strong law of large numbers

Theorem 6.9 (SLLN)**.** *Suppose that $X_1, \ldots, X_n$ are iid. Then $\bar{S}_n \xrightarrow{a.s.} \mu$ as $n \to \infty$ if and only if $E[|X_1|] < \infty$ and $E[|X_1|] = \mu$.*

Remark 6.10 (Finite Expectations and Forward Direction Proof Steps)**.** Recall that

$$E[|X|] < \infty \iff \sum_{n\geq 1} \mathbb{P}[|X| > n] < \infty.$$

This is because we know that

$$E[|X|] = \int_0^\infty \mathbb{P}[|X| \geq x]dx.$$

And as we can see by taking unit-width rectangles, i.e., to show that

$$f(1) + f(2) + \cdots \leq \int_0^\infty f(x)dx \leq f(0) + f(1) + f(2) + \cdots,$$

we have that when f is non-increasing (for all large enough $x \gg 1$):

$$\int_0^\infty f(x)dx < \infty \iff \sum_{n\geq 1} f(n) < \infty.$$

Now, secondly, we also observe that if $\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} \mu$, then $\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} \xrightarrow{a.s.} 0$. To see this we look at the limiting behavior of $\frac{X_n}{n} = \frac{S_n}{n} - \left(\frac{n-1}{n}\right)\frac{S_{n-1}}{n-1} \longrightarrow 0$. So we conclude that $\frac{X_n}{n} \xrightarrow{a.s.} 0$.
Claim: This observation implies that X_1 is integrable.

Proof. Suppose that in fact X_1 is not integrable. Then we have that $\sum_{n\geq 1} \mathbb{P}[|X_1| > n] = +\infty$. So in fact since the X_i's are independent, by Borel-Cantelli (II), $\mathbb{P}\left[\limsup_{n\to\infty}\{|X_n| > n\}\right] = 1$. But this contradicts our fact proven above that $\frac{X_n}{n} \to 0 \neq 1$. □

Finally, in summary if $E[|X_1|] < \infty$, we have proved that $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu$, with $\mu = E[|X_1|]$. Also, though we have that $\frac{S_n}{n} \xrightarrow{a.s.} \mu_0$, so we obtain that in fact this $\mu_0 = E[|X_1|]$. This concludes the proof of the forward direction of the SLLN.

Lemma 6.11 (Real Analysis Fact)**.** *If $\sum_{n\geq 1} \frac{x_n}{n} < \infty$, then immediately $\frac{x_1 + \cdots + x_n}{n} \to 0$.*

Proof. By hypothesis, we have that $S_n \xrightarrow{S}$ as $n \to \infty$. Now if $S_n := \sum_{i=1}^n \frac{x_i}{i}$, then $x_n = (S_n - S_{n-1}) \cdot n$. So

$$\begin{aligned}\frac{x_1 + \cdots + x_n}{n} &= \frac{1}{n}\left(S_1 - S_0 + 2(S_2 - S_2) + \cdots + n(S_n - S_{n-1})\right) \\ &= \frac{1}{n}\left(S_0 - S_1 - S_2 - \cdots - S_{n-1} + nS_n\right) \\ &= S_n - \frac{S_0 + \cdots + S_{n-1}}{n} \longrightarrow S := 0. \qquad \square\end{aligned}$$

Theorem 6.12 (Kolmogorov)**.** *If Y_n is a sequence of independent random variables such that $\sum_{n\geq 1} \operatorname{Var}(Y_n) < \infty$, then $\sum_{n\geq 1} Y_n$ is a.s. convergent. If $Y_n := \frac{X_n}{n}$ satisfies $\sum_{n\geq 1} \frac{\operatorname{Var}(X_n)}{n^2} < \infty$, then $\sum_{n\geq 1} \frac{X_n - E[X_n]}{n} < \infty$. As a consequence, if $X_1, \cdots, X_n$ are iid with mean μ and finite variance, then $S_n \xrightarrow{a.s.} \mu$.*

As a comment, if $\operatorname{Var}(X_1) < \infty$, we have already proved that

$$\mathbb{P}\left[|S_n - \mu| > \varepsilon\right] \leq \frac{\operatorname{Var}(X_1)}{n\varepsilon^2},$$

but in this case it's NOT Borel-Cantelli that provides the a.s. convergence.

Remark 6.13. To finish the sketch of the reverse direction of the SLLN, we next sketch the last remaining key ideas to this part of the proof. In particular, we can 1) use Borel-Cantelli to show the following:

$$\sum_{n\geq 1} \mathbb{P}[Y_n \neq X_n] = \sum_{n\geq 1} \mathbb{P}[|X_n| > n] = \sum_{n\geq 1} \mathbb{P}[|X_1| > n] \implies \mathbb{P}[\limsup_{n\to\infty}\{X_n \neq Y_n\}] = 0,$$

i.e., for all large enough n, $X_n = Y_n$ (a.s.) – so that

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} \mu \iff \frac{Y_1 + \cdots + Y_n}{n} \xrightarrow{a.s.} \mu.$$

Then 2) we can use Kolomogorov's theorem on the variance to show that $\frac{Y_1 + \cdots + Y_n}{n} \xrightarrow{a.s.} \mu$.

Corollary 6.14 (Consequences of the SLLN)**.** *If $X_1, \ldots, X_n$ are iid, then for any $f : \mathbb{R} \to \mathbb{R}$ which is measurable and bounded, we have that $\sum_{i=1}^n \frac{f(X_i)}{n} \to E[f(X_i)]$. For example, suppose that $f(x) := \chi_{[a,b]}(x)$. Then $\frac{1}{n}\sum_{i=1}^n \chi_{[a,b]}(X_i) \to E[\chi_{[a,b]}(X_1)] = \mathbb{P}[a \leq X_1 \leq b]$, where the former (non-limiting) sum corresponds to the relative frequence of X_i in $[a, b]$. Now for the special case where $a = -\infty$ and $b = x$, we obtain that*

$$\frac{1}{n}\sum_{i=1}^n \chi_{(-\infty,x]}(X_i) \longrightarrow F_{X_1}(x) \equiv \mathbb{P}[X_1 \leq x].$$

Example 6.15 (Spring 2018, #2)**.** Suppose that f is a continuous function on $[0, 1]$. Use the Law of Large Numbers to prove that

$$\lim_{n\to\infty} \int_0^1 \cdots \int_0^1 f\left((x_1 \cdots x_n)^{1/n}\right) dx_1 \cdots dx_n = f\left(\frac{1}{e}\right).$$

Example 6.16 (Fall 2018, #1)**.** Use the SLLN to find the following limit:

$$\lim_{n\to\infty} \int_0^1 \cdots \int_0^1 \frac{x_1^2 + \cdots + x_n^2}{x_1 + \cdots + x_n} dx_1 \cdots dx_n.$$

Example 6.17 (Spring 2017, #3)**.** Let $X_1, X_2, \ldots$ be iid random variables uniformly distributed on $[0,1]$. Show that with probability 1,

$$\lim_{n\to\infty} (X_1 \cdots X_n)^{1/n},$$

exists and compute its value.

Example 6.18 (Spring 2015, #8)**.** Let $(X_n)_{n\geq 1}$ be an iid sequence of positive random variables such that $E[X_1] < \infty$. Let

$$N_t := \sup\{n : X_1 + \cdots + X_n \leq t\}.$$

Show that

$$\frac{N_t}{t} \xrightarrow[n\to\infty]{a.s.} \frac{1}{E[X_1]},$$

where the convergence is in the almost sure sense.

6.4 Concentration inequalities

Theorem 6.19. *If $X_1, \ldots, X_n$ are iid such that $a \leq X_i \leq b$ for all i, then*

$$\mathbb{P}(|\bar{S}_n - \mu| > \varepsilon) \leq 2e^{-2n\varepsilon^2/(b-a)^2}.$$

Proof Notes. We require the following lemma: If $a \leq X \leq b$, then $\mathrm{Var}(X) \leq \frac{(b-a)^2}{4}$. □

Proposition 6.20. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz if*

$$|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \leq \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

Suppose that $X_1, \ldots, X_n$ are iid, $X_1 \sim N(0,1)$, and that f is 1-Lipschitz on $\mathbb{R}^n$ such that f is differentiable and $\sup_{x\in\mathbb{R}^n} |\nabla f(x)| \leq 1$. Then

$$\mathbb{P}\left(|f(x_1, \ldots, x_n) - E[f(x_1, \ldots, x_n)]| > \varepsilon\right) \leq e^{-\varepsilon^2/2}.$$

Example 6.21 (Cauchy-Schwarz and 1-Lipschitz Functions)**.** An example of a 1-Lipschitz function f is given by $f(x_1, \ldots, x_n) := \frac{x_1 + \cdots + x_n}{\sqrt{n}}$. To show this, we can pull a trick out of our hat in the form of "forcing" the Cauchy-Schwarz inequality in the following form:

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i\right)^2 \leq \frac{1}{n}\left(\sum_{i=1}^n 1^2\right)\left(\sum_{i=1}^n a_n^2\right) = \sum_{i=1}^n a_i^2.$$

In our case, we have that

$$|f(\vec{x}) - f(\vec{y})| \leq \sum_{i=1}^n |x_i - y_i| \implies \frac{1}{\sqrt{n}} \sum_{i=1}^n \leq \sqrt{\sum_{i=1}^n |x_i - y_i|^2}.$$

Example 6.22. If $X_1, \ldots, X_n$ are iid with $X_1 \sim N(0,1)$, we get that for $f(x_1, \ldots, x_n) := \frac{1}{\sqrt{n}}(x_1 + \cdots + x_n)$:

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}} \sum_i X_i - E\left[\frac{1}{\sqrt{n}} \sum_i X_i\right]\right| > \lambda\right) \leq 2e^{-\lambda^2/2}$$

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| > \sqrt{n}\lambda\right) \leq 2e^{-\lambda^2/2}, \ \forall \lambda > 0.$$

And if we take $\lambda := \sqrt{n}\varepsilon$, then

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n} X_i\right| > \varepsilon\right) \leq 2e^{-n\varepsilon^2/2}.$$

Example 6.23 (Spring 2018, #6)**.** Let $(X_n)_{n\geq 1}$ be a sequence of iid random variables with

$$\mathbb{P}\left(X_n = 1\right) = \frac{1}{2} = \mathbb{P}\left(X_n = -1\right).$$

Let $(Y_n)_{n\geq 1}$ be a bounded sequence of random variables such that $\mathbb{P}\left(Y_n \neq X_n\right) \leq e^{-n}$. Show that

$$\frac{1}{n}E[(Y_1 + \cdots + Y_n)^2] \longrightarrow 1,$$

as $n \to \infty$.

Example 6.24 (Spring 2017, #2; Spring 2016, #1)**.** Let X be a random variable with mean zero and finite variance σ^2. Prove that for every $c > 0$,

$$\mathbb{P}\left(X > c\right) \leq \frac{\sigma^2}{\sigma^2 + c^2}.$$

(HINT: Combine the inequality $E[c - X] \leq E[(c - X)\chi_{X<c}]$ with the Cauchy-Schwarz inequality.)
(HINT: Write $c - X = (c - X)_+ - (c - X)_-$ and then use Cauchy-Schwarz.)

7 Normal distributions

7.1 One-dimensional case

Definition 7.1 (Normal and Standard Normal Distrubutions)**.** We say that $Z \sim N(0,1)$ and $X \sim N(\mu, \sigma^2)$ if their probability density functions satisfy

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},\ x \in \mathbb{R}$$
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2},\ x \in \mathbb{R}.$$

One key important fact relating these two normal distributions is that if $X \sim N(\mu, \sigma^2)$, then the random variable $Z := \frac{X-\mu}{\sigma} \sim N(0,1)$. In particular, given this definition, $F_Z(z) = F_X(\sigma z + \mu)$ and $f_Z(z) = \sigma \cdot f_X(\sigma z + \mu)$.

Proposition 7.2 (Properties of the Standard Normal)**.** *We have the following properties of a standard normal random variable $Z \sim N(0,1)$:*

(1) $E[Z] = 0$ and $\mathrm{Var}(Z) = 1$;

(2) $M_Z(\lambda) = E[e^{\lambda Z}] = e^{\lambda^2/2}$ for all $\lambda \in \mathbb{C}$.

Proof of (1). First, we compute that

$$E[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz = \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz = 0,$$

since the right-hand-side integral is of an odd function over a symmetric interval. Next, we can compute by repeated integration by parts that

$$\begin{aligned}
\mathrm{Var}(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \left(-e^{-z^2/2}\right) dz \\
&= z e^{-z^2/2} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz \\
&= 1.
\end{aligned}$$

□

Proof of (2). We have that

$$\begin{aligned}
M_Z(\lambda) = E[e^{\lambda Z}] &= \int_{-\infty}^{\infty} \frac{e^{\lambda z}}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= \frac{e^{\lambda^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2\lambda z + \lambda^2)} dz \\
&= e^{\lambda^2/2}.
\end{aligned}$$

□

Corollary 7.3. *If $X \sim N(\mu, \sigma^2)$, then its moment generating function is given by $M_X(\lambda) = e^{\lambda\mu + (\lambda\sigma)^2/2}$.*

Proof. We can write $X = \mu + \sigma Z$, where $Z \sim N(0,1)$. This implies that

$$M_X(\lambda) = E[e^{\lambda X}] = E[e^{\lambda(\mu + \sigma Z)}] = e^{\lambda\mu} E[e^{\lambda\sigma Z}] = e^{\lambda\mu + \lambda^2\sigma^2/2}.$$

□

Example 7.4 (Fall 2018, #3)**.** Let $(Z_n)_{n\geq 1}$ be iid standard normal random variables, and let $(a_n)_{n\geq 1} \subset \mathbb{R}$. Prove that $\sum_{n\geq 1} a_n Z_n^2 < \infty \iff \sum_{n\geq 1} a_n < \infty$.

Example 7.5 (Spring 2016, #6)**.** Let $X_1, X_2, \ldots$ be iid standard normal random variables, and for $x \in (-1, 1)$, set

$$Y := \sum_{n\geq 1} x^n X_n.$$

Show that the sum defining Y converges and find its distribution.

7.2 Multidimensional normal distributions

Definition 7.6. The random variable $Z := (Z_1, \ldots, Z_d)$ is a *standard multidimensional normal vector* if $Z_1, \ldots, Z_d$ are iid with $Z_1 \sim N(0, 1)$. We write that $X = (X_1, \ldots, X_d) \sim N(\mu, C)$ with $\mu \in \mathbb{R}^d$ and covariance matrix C if [1]

$$f_X(x_1, \ldots, x_d) = \frac{1}{\sqrt{(2\pi)^d (\det C)^{1/2}}} e^{-\langle C^{-1}(x-\mu), x-\mu\rangle/2}.$$

One particularly nice property relating these two types of multidimensional normal variables is that $X \sim N(\mu, C) \iff Z = C^{-1/2}(X - \mu) \sim N(0, \mathrm{Id}) \iff X = \mu + C^{1/2} Z$ with $Z \sim N(0, \mathrm{Id})$.

Exercise 7.7. Show that if $X = (X_1, \ldots, X_n) \sim N(\mu, C)$, then $C_{ij} = \mathrm{Cov}(X_i, X_j)$ and $E[X_i] = \mu_i$.

Proposition 7.8 (Properties and Transformations)**.** *We have the following properties and transformation results relating multidimensional normal vectors:*

(1) $X \sim N(\mu, C) \iff X = \mu + C^{1/2} Z$ *where* $Z = (Z_1, \ldots, Z_d)$ *for* Z_i *iid and* $Z_1 \sim N(0, 1)$.

(2) If $X \sim N(\mu, C)$, $A : \mathbb{R}^d \to \mathbb{R}^m$ *is a linear map, then* $Y := AX \sim N(A\mu, ACA^T)$.

(3) If $X \sim N(\mu, C)$, *then for any* $a_1, \ldots, a_d \in \mathbb{R}$, $\sum_{i=1}^d a_i X_i \sim N\left(\sum_{i=1}^d a_i u_i, \sum_{i,j=1}^n C_{ij} a_i a_j\right)$.

(4) If $Z \sim N(0, \mathrm{Id})$ *and* U *is an orthogonal matrix, then* $\widetilde{Z} := UZ \sim N(0, \mathrm{Id})$.

Sketch of (2). We first cite the following theorem: If $M_X(\lambda) = M_Y(\lambda)$ for all $||\lambda||_2 < \varepsilon$ $(\forall \varepsilon > 0)$, then $X \sim Y$ have the same multidimensional distrubution.

Step 1: Notice that

$$M_X(\lambda) = E[e^{\langle \lambda, X\rangle}] = E\left[e^{\sum_i \lambda_i X_i}\right].$$

Step 2: We compute that

$$\begin{aligned} M_{AX}(\lambda) &= E\left[e^{\langle \lambda, AX\rangle_{\mathbb{R}^m}}\right] = E\left[e^{\langle A^T\lambda, X\rangle_{\mathbb{R}^d}}\right] \\ &= e^{\langle A^T\lambda, \mu\rangle + \langle CA^T\lambda, A^T\lambda\rangle} \\ &= e^{\langle \lambda, A\mu\rangle + \langle ACA^T\lambda, \lambda\rangle} \\ &= M_{N(A\mu, ACA^T)}(\lambda). \end{aligned}$$

□

Example 7.9. Let $X := [X_1, X_2]^T \sim N(\mu, C)$ where $X = \mu + C^{1/2} Z$ for $Z \sim N(0, \mathrm{Id})$. Suppose that

$$C = \begin{bmatrix} \mathrm{Cov}_{X_1} & 0 \\ 0 & \mathrm{Cov}_{X_2} \end{bmatrix},$$

[1] Note that if $C \in \mathbb{C}^{d\times d}$ is symmetric, $C \geq 0$, then $C = UDU^T$ where U is orthogonal and D is a diagonal matrix of eigenvalues of C. In particular, this shows that $C^{1/2} = UD^{1/2}U^T$ in the computations from the last definitions.

where $\mathrm{Cov}_{X_1,X_2} = 0 \implies X_1, X_2$ are normal and independent. Then

$$X = \mu + C^{1/2} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \mu + \begin{bmatrix} \mathrm{Cov}_{X_1}^{1/2} Z_1 \\ \mathrm{Cov}_{X_2}^{1/2} Z_2 \end{bmatrix},$$

so that $X_i = \mu_i \, \mathrm{Cov}_{X_i}^{1/2} Z_i$ are independent with the Z_i independent.

Theorem 7.10 (Concentration Inequality)**.** *If $Z_1, \ldots, Z_d$ are iid with $Z_1 \sim N(0,1)$, then for any 1-Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$, we have that*

$$\mathbb{P}\left(|f(Z_1, \ldots, Z_d) - E[f(Z_1, \ldots, Z_d)]|\right) \leq C_1 \cdot e^{-C_2 \varepsilon^2},$$

where the sharp constant is given by $C_2 = 1/2$.

Proof Components. We cite *Jensen's inequality* which states that for ϕ convex: $E[\phi(X)] \geq \phi(E[X])$. We also use the identity that $E[e^{\lambda f(Z)}]E[e^{-\lambda f(Z)}] = E[e^{\lambda(f(Z_1) - f(Z_2))}]$. □

Example 7.11 (Fall 2016, #4)**.** Let (X, Y) be a normal vector in $\mathbb{R}^2$ with mean zero and covariance matrix:

$$\Sigma = \begin{bmatrix} 5 & 1 \\ 1 & 10 \end{bmatrix}.$$

Find $E[X^2 Y^2]$.

Example 7.12 (Spring 2016, #2)**.** Let $X = (X_1, X_2)$ be a Gaussian vector with zero mean and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where $|\rho| < 1$. Find a matrix A such that $X = AZ$ where Z is a standard normal vector, and derive the characteristic function of X as a function of ρ.

Example 7.13 (Spring 2015, #3)**.** Assume that (X, Y) is a joint normal vector with $E[X] = E[Y] = 0$. Show that

$$E[X^2 Y^2] \geq E[X^2] E[Y^2],$$

with equality if and only if X, Y are independent.

8 Central limit theorem

Theorem 8.1 (Central Limit Theorem)**.** *Let $X_1, \dots, X_n, \dots$ be independent iid random variables with mean μ and finite variance. If $-\infty < a < b < \infty$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Example 8.2 (Spring 2018, #4)**.** Let $X_1, \dots, X_n$ be iid random variables with mean μ and variance $\sigma^2 < \infty$. Let f be a function continuously differentiable at the point μ. Prove that the sequence of random variables

$$\sqrt{n}\left(f\left(\frac{X_1 + \cdots + X_n}{n}\right) - f(\mu)\right),$$

converges in distribution to a normal random variable. What is the mean and the variance of the limit?

Proof. If we define $Y_n := \sqrt{n}\left(\frac{X_1+\cdots+X_n}{n} - \mu\right)$, then $Y_n \xrightarrow{\text{dist}} Y$, where $Y \sim N(0, \sigma^2)$ (by the Central Limit Theorem). Also, by the SLLN $Y_n/\sqrt{n} \xrightarrow{\mathbb{P}} 0$. □

Example 8.3 (Spring 2018, #5)**.** Let $X_1, \dots, X_n, \dots$ be iid random variables with $E[X_1] = 0$ and $\mathrm{Var}(X_1) = 1$. Define $S_n := X_1 + \cdots + X_n$. Prove that

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = +\infty.$$

Example 8.4 (Spring 2017, #6)**.** Let (X_i) be iid random variables which are uniformly distributed on $[0, 2]$. Let $S_n := X_1 + \cdots + X_n$. Show that

$$\frac{3\sqrt{3}}{2} n^{1/6}\left(\sqrt[3]{S_n} - \sqrt[3]{n}\right) \implies {}^w Z,$$

where Z is a standard normal random variable.

Example 8.5 (Spring 2016, #7)**.** Let $X_1, X_2, \dots$ be independent random variables such that $X_n \sim$ Bernoulli(p_n) for some $p_n > 0$. Show that if $np_n(1 - p_n) \to \infty$, then

$$\frac{X_n - np_n}{\sqrt{np_n(1 - p_n)}} \implies {}^w N(0, 1).$$

Example 8.6 (Spring 2015, #2)**.** Assume that $X_1, X_2, \dots$ is a sequence of iid random variables such that for some $\alpha < 1/2$,

$$\frac{X_1 + \cdots + X_n}{n^\alpha} \xrightarrow[n\to\infty]{a.s.} m,$$

for some real number m. Show that almost surely $X_i = 0$.

9 Fall 2018 Probability I final exam solutions

9.1 Problem 1

Problem Statement 9.1. Show that a random variable X such that

$$E[e^{\lambda X}] \le e^{2|\lambda|^3}, \ \forall \lambda \in [-1,1],$$

satisfies $X = 0$ almost surely.

Proof. Let $A_n := \{\omega : \lambda |X|(\omega) > \frac{1}{n}\}$. We can show that the sum $\sum_{n\ge1} \mathbb{P}(A_n) < \infty$, by evaluating the corresponding integral

Maybe $P(A_n) \le \frac{e^{2|\lambda|^3}}{e^{1/n}}$

$$\int_1^\infty \mathbb{P}\left(\lambda|X| > \frac{1}{t}\right) dt = \int_1^\infty \mathbb{P}\left(e^{\lambda|X|} > e^{\frac{1}{t}}\right) dt$$
$$= \int_1^e \mathbb{P}\left(e^{\lambda|X|} > t\right) dt \le 2\int_1^e \mathbb{P}\left(e^{\lambda X} > t\right) dt$$
$$\le 2E[e^{\lambda X}] \le 2e^{2|\lambda|^3} < \infty.$$

Here you have to use a substitution $e^{\frac{1}{t}} = s$ so $t = \frac{1}{\ln s}$, $dt = -\frac{1}{s(\ln s)^2}$ and then things get stuck here.

So that the sum converges. By Borel-Cantelli we have that

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0,$$

and since we can take $\lambda = 1$ and still guarantee convergence of the infinite sum, this implies that

$$\mathbb{P}\left(\limsup_{n\to\infty}\{|X| > \frac{1}{n}\}\right) = 0,$$

which guarantees that $X = 0$ almost surely. $\square$

You can try here that $E(X) = \lim_{\lambda\to 0} \frac{E[e^{\lambda X}]-1}{\lambda} \le 0$ and also $E(-X) = \lim_{\lambda\to 0} \frac{E[e^{-\lambda X}]-1}{\lambda} \le 0$ so Thus $E(X)=0$. Similarly $2E[X^2] = \lim_{\lambda\to 0} \frac{E(e^{\lambda X}) - \lambda E(X) - 1}{\lambda^2} \le 0$

9.2 Problem 2

Problem Statement 9.2. Give an example of random variables X_n which converge to 0 almost surely, but such that $\sum_{n\ge1} \mathbb{P}(|X_n| \ge \varepsilon) = \infty$, for any $\varepsilon > 0$.

Solution. For $n \ge 1$, let $X_n \sim \text{Uniform}([0,1/n])$. Then for any $\omega \in \Omega$,

If this is so

$$1 = \mathbb{P}(X_n(\omega) \ne 0) = \mathbb{P}(\omega \in [0,1/n]) \longrightarrow 0,$$

as $n \to \infty$. For any $\varepsilon > 0$, we can also estimate

$$\mathbb{P}(|X_n| > \varepsilon) = \int_0^{\min(\varepsilon,1/n)} \frac{1}{1/n - 0} dt = n\min(\varepsilon, 1/n) \implies \sum_{n\ge1} \mathbb{P}(|X_n| > \varepsilon) = +\infty. \qquad \square$$

You probably want here $X_n = n \mathbb{1}_{[0,1/n]}(U)$ where U is uniform on $[0,1]$.

9.3 Problem 3

Problem Statement 9.3. The moment generating function of a random variable X is defined to be $M_X(\lambda) = E[e^{\lambda X}]$, which may only be convergent for some $\lambda \in \mathbb{R}$. Prove the following:

a. If $M_X(\lambda)$ is defined for some $0 < |\lambda| < \delta$, then

$$\mathbb{P}(|X| \ge x) \le e^{-cx}\left(M_X(c) + M_X(-c)\right),$$

for any $x > 0$ and $\delta > c > 0$;

$P(|X| \ge x) = P(X > x) + \underbrace{P(X < -x)}_{P(-X > x)} \le e^{-cx} E[e^{cX}] + e^{-cx} E[e^{-cX}]$

b. Show that if for a given random variable X there exist constants $c, C > 0$ such that $\mathbb{P}(|X| \geq x) \leq Ce^{-cx}$ for any $x > 0$, then $E[e^{\lambda X}] < \infty$ for $|\lambda| < c$;

c. If $M_X(\lambda)$ exists for $\lambda < \delta$, with $\delta > 0$, then for all $\alpha > 0$, $E[|X|^\alpha] < \infty$. Justify that

$$M_X(\lambda) = \sum_{k \geq 0} \frac{E[X^k]\lambda^k}{k!}.$$ for small λ.

We also have that

$$\frac{d^{(k)}}{d\lambda^{(k)}} M_X(\lambda)\Big|_{\lambda=0} = E[X^k].$$

Proof of (a). I am still unclear on how to prove this inequality. □

Proof of (b). Here use $E[e^{\lambda|X|}] = 1 + \lambda\int_0^\infty e^{\lambda x}\,\mathbb{P}(|X|>x)\,dx < \infty$ if $\lambda < c$ □

Proof of (c). key: $\frac{c^{2n}x^{2n}}{(2n)!} \leq e^{c|x|}$. Then differentiate both sides of (*) □

9.4 Problem 4

Problem Statement 9.4. Consider and prove each of the statements below:

(a) Assume X, Y are two bounded random variables. If for any integer numbers $m, n \geq 0$, we have

$$E[X^m Y^n] = E[X^m]E[Y^n],$$

show that X, Y must be independent.

(b) Take random variables X, Y such that for some constant $c > 0$, we have that

$$\mathbb{P}(|X| \geq x) + \mathbb{P}(|Y| \geq x) \leq e^{-cx}, \ \forall x > 0.$$

Precisely, we need to show that for any integers $m, n \geq 0$, $X^m Y^n$ is integrable. Moreover, we show that if for any such $m, n \geq 0$,

$$E[X^m Y^n] = E[X^m]E[Y^n],$$

then X, Y are necessarily independent.

(c) Assume that X, Y are independent random variables such that their moment generating functions, $M_X(\alpha) = E[e^{\alpha X}]$ and $M_Y(\beta) = E[e^{\beta Y}]$, are finite for all $0 < |\alpha|, |\beta| < \delta$. Prove that if

$$E[e^{\alpha X + \beta Y}] = E[e^{\alpha X}]E[e^{\beta Y}], \forall |\alpha|, |\beta| < \delta,$$

then X, Y must be independent.

Proof of (a). Key: use the fact that any continuous function on a compact interval can be approximated by polynomials. From continuous functions you can extend to $f(X)g(Y)$ where $f(t) = \mathbb{1}_{(-\infty,x]}(t)$, $g(t) = \mathbb{1}_{(-\infty,y]}(t)$. □

Proof of (b). We first assume that X, Y are independent for the first part of the proof. Notice that by trading expansions with inclusion-exclusion, we have that for any $x > 0$:

$$\begin{aligned}
e^{-cx} &\geq \mathbb{P}(|X| \geq x) + \mathbb{P}(|Y| \geq x) \\
&= \mathbb{P}(|X| \geq x \vee |Y| \geq x) - \mathbb{P}(|X|^m \geq x^m)\,\mathbb{P}(|Y|^n \geq x^m) \\
&\geq \mathbb{P}(|X| \geq x \vee |Y| \geq x) - \frac{E[|X|^m]E[|Y|^n]}{x^{m+n}}
\end{aligned}$$

$$
\begin{aligned}
&= 1 - \mathbb{P}\left(|X| \leq x \wedge |Y| \leq x\right) - \frac{E[|X|^m]E[|Y|^n]}{x^{m+n}} \\
&= \mathbb{P}\left(|X|^m|Y|^n \geq x^{m+n}\right) - \frac{E[|X|^m]E[|Y|^n]}{x^{m+n}}.
\end{aligned}
$$

Now, X^mY^n is integrable whenever its expectation is finite. So it suffices to show that

$$
\begin{aligned}
\int_1^\infty \mathbb{P}\left(X^mY^n \geq t^{m+n}\right) dt &\leq \int_1^\infty e^{-cx}dx + E[|X|^m]E[|Y|^n] \int_1^\infty \frac{dx}{x^{m+n}} \\
&= \frac{1}{c}e^{-c} - \frac{E[|X|^m]E[|Y|^n]}{(m+n-1)} < \infty.
\end{aligned}
$$

For the second statement, notice that the conclusion is trivial if $m = 0$ or $n = 0$ since $E[1] = 1$. Thus we prove the following lemma for integral powers $m, n \geq 1$:

Lemma 9.5. *Two random variables X, Y are independent if and only if X^m, Y^n are independent for all natural numbers $m, n \geq 1$.*

Proof. Suppose that we have

$$
\begin{aligned}
\mathbb{P}\left(X^m \geq x_1 \wedge Y^n \geq x_2\right) &= \mathbb{P}\left(X \geq x_1^{1/m} \wedge Y \geq x_2^{1/n}\right) \\
&= \mathbb{P}\left(X \geq x_1^{1/m}\right)\mathbb{P}\left(Y \geq x_2^{1/n}\right) \\
&\iff X, Y \text{ are independent.}
\end{aligned}
$$

□

□

Proof of (c). □

9.5 Problem 5

Problem Statement 9.6. For a sequence $(X_n)_{n\geq 1}$ of exponential random variables, say with $X_i \sim \mathrm{Exp}(\lambda_i)$ for $\lambda_i > 0$ $\forall i \geq 1$, find sequences of constants a_n, b_n, c_n, d_n such that the two sequences

$$
\begin{aligned}
Y_n &:= a_n \max(X_1, \ldots, X_n) + b_n \\
Z_n &:= c_n \min(X_1, \ldots, X_n) + d_n,
\end{aligned}
$$

both converge (weakly) to some limiting distribution. Identify the limiting distributions.

Solution for the Min-Variable Case. Notice that by independence, we must have the product

$$
\begin{aligned}
\mathbb{P}\left(\min(X_1, \ldots, X_n) > \frac{y - d_n}{c_n}\right) &= \mathbb{P}\left(X_i > \frac{y - d_n}{c_n},\ \forall 1 \leq i \leq n\right) \\
&= \prod_{i=1}^n \mathbb{P}\left(X_i > \frac{y - d_n}{c_n}\right) = \prod_{i=1}^n \exp\left[-\lambda_i\left(\frac{y - d_n}{c_n}\right)\right] \\
&= e^{\frac{d_n}{c_n}(\lambda_1 + \cdots + \lambda_n)} \times \exp\left[-y\frac{(\lambda_1 + \cdots + \lambda_n)}{c_n}\right],
\end{aligned}
$$

which means that $\frac{1}{c_n}(Z_n - d_n)$ is a multiple of an exponentially distributed random variable with parameter given by $\frac{(\lambda_1 + \cdots + \lambda_n)}{c_n}$. Now to remove the inconvenient multiple, and to simplify the problem, we might as well set the two sequences to be constant as $(c_n, d_n) := (1, 0)$ for all $n \geq 1$. Then if we further make the restriction that for

$$\lambda := \sum_{i \geq 1} \lambda_i, \; 0 < \lambda < \infty,$$

we obtain a limiting distribution for Z_n which is exponential with parameter λ. □

Solution for the Max-Variable Case. For the purposes of simplicity, we assume that all of the parameters $\lambda_i = \lambda$ are identical here (otherwise for the observations part of the question, the limiting distribution gets extremely hairly). For this exploratory analysis, we will use independence coupled with an application of inclusion-exclusion (binomial theorem) to find that

$$\begin{aligned}
\mathbb{P}\left(\max(X_1, \ldots, X_n) > \frac{y - b_n}{a_n}\right) &= \mathbb{P}\left(X_1 > \frac{y - b_n}{a_n} \vee \cdots \vee X_n > \frac{y - b_n}{a_n}\right) \\
&= \sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} \mathbb{P}\left(X_1 > \frac{y - b_n}{a_n}\right)^k \\
&= 1 - \left(1 - e^{-\lambda\left(\frac{y - b_n}{a_n}\right)}\right)^n.
\end{aligned}$$

Now let's set $a_n \equiv 1$ and declare that $b_n = \log(1/n)$, i.e., so that we get an exponential function in the limiting case of the last equation:

$$\mathbb{P}\left(\max(X_1, \ldots, X_n) > \frac{y - b_n}{a_n}\right) = 1 - \left(1 - \frac{e^{-\lambda y}}{n^\lambda}\right)^n \longrightarrow 1 - \exp\left[-\frac{1}{\lambda} e^{-\lambda y}\right].$$

Now by fiddling with the probability equations for the cumulative density, we can find that

$$\mathbb{P}\left(\frac{1}{\lambda} \log(1/Y_n) \geq y\right) \xrightarrow[n \to \infty]{w} e^{-y/\lambda}.$$

So in summary, the sequence of random variables defined by $\log(Y_n^{-1/\lambda}) \to \operatorname{Exp}(1/\lambda)$ tend to an exponentially distributed limit as $n \to \infty$. ✓ Good! □

9.6 Problem 6

Problem Statement 9.7. Assume that $(X_n)_{n \geq 1}$ are iid with $X_1 \sim N(0, 1)$. Show that for any $\lambda > 1/2$,

$$\bar{S}_n = \frac{1}{n^\lambda} \sum_{i=1}^{n} X_i \xrightarrow[n \to \infty]{a.s.} 0.$$

Proof. The partial sums $S_n := X_1 + \cdots + X_n$ satisfy $S_n \sim N(0, n)$, with corresponding distribution function $f_{S_n}(x) = \frac{1}{\sqrt{2\pi n}} e^{-x^2/2n}$. Now let $\varepsilon > 0$ and consider the convergence of the sums

$$\sum_{n \geq 1} \mathbb{P}\left(\frac{|S_n|}{n^\lambda} > \varepsilon\right) < \infty \iff I_\varepsilon := \int_1^\infty \mathbb{P}\left(|S_t| > t^\lambda \varepsilon\right) dt < \infty.$$

With the change of variables $s = x/\sqrt{2t}$ and $\sqrt{2t} \cdot ds = dx$, we can re-write the last integral as the double integral:

$$I_\varepsilon = \int_1^\infty \int_0^{\varepsilon t^\lambda} \frac{1}{\sqrt{\pi}} e^{-s^2} ds dt.$$

A simpler proof is using the large deviation bound

$P(|\bar{S}_n| > \varepsilon) \leq 2e^{-\varepsilon^2/2n}$.

Since the density function $f_{S_t}(x)$ is non-negative, continuous (hence measurable) and integrable on $[0,\infty)$, Fubini-Tonelli tells us that we can swap the order of integration in the last equation where we also use that $\lambda > 1/2 \implies \frac{1}{\lambda} < 2$:

$$\begin{aligned}
I_\varepsilon &= \int_1^\infty \int_0^\infty \frac{e^{-s^2}}{\sqrt{\pi}} \chi_{[0,\varepsilon t^\lambda]}(s)\,ds\,dt \\
&= \int_0^\infty \int_\varepsilon^{(s/\varepsilon)^{1/\lambda}} \frac{e^{-s^2}}{\sqrt{\pi}}\,dt\,ds \\
&= \int_0^\infty \frac{e^{-s^2}}{\sqrt{\pi}} \left[\left(\frac{s}{\varepsilon}\right)^{1/\lambda} - \varepsilon\right] ds \\
&\le \int_0^\infty \frac{s^2 e^{-s^2}}{\sqrt{\pi}\varepsilon^{1/\lambda}} - \frac{\varepsilon}{2} \\
&= \frac{1}{4\varepsilon^{1/\lambda}} - \frac{\varepsilon}{2} < \infty.
\end{aligned}$$

So we obtain convergence of the infinite sum we wished to bound above. And then by Borel-Cantelli, we have that

$$\mathbb{P}\left(\limsup_{n\to\infty}\{|S_n| n^{-\lambda} > \varepsilon\}\right) = 0,$$

which as we have argued before implies almost sure convergence of S_n/n^λ to zero as $n \to \infty$. □

9.7 Problem 7

I think the best way to do this is to repeat the proof of the strong LLN with the appropriate changes.

Problem Statement 9.8. Show that for any sequence $(X_n)_{n\ge1}$ of iid random variables,

$$\frac{1}{n^2}\sum_{i=1}^n iX_i \xrightarrow[n\to\infty]{a.s.} \frac{m}{2} \iff X_1 \text{ is integrable and } E[X_1] = m.$$

Proof. Let $X_1, \dots, X_n, \dots$ be any sequence of iid random variables. Set $R_n := \sum_{i=1}^n iX_i$ and $S_n := X_1 + \cdots + X \cancel{= n}$ n as usual. First, we can compute a more useful formula for R_n using summation by parts:

$$\begin{aligned}
R_n &= n \cdot S_n - \sum_{i=1}^{n-1} S_i(i+1-i) = n \cdot S_n - \sum_{i=1}^{n-1} S_i \\
&= n \cdot S_n - \left[(n-1)X_1 + (n-1)X_2 + \cdots + 2X_{n-2} + X_{n-1}\right].
\end{aligned}$$

Then the scaled sums are expanded by

$$\begin{aligned}
\frac{R_n}{n^2} &= \frac{S_n}{n} - \sum_{k=1}^{n-1} \frac{(n-k)}{n^2} X_k \\
&= \bar{S}_n - \frac{S_{n-1}}{n} + \frac{R_{n-1}}{n^2} \\
&= \bar{S}_n - \frac{(n-1)}{n}\bar{S}_{n-1} + \frac{(n-1)^2}{n^2}\frac{R_{n-1}}{(n-1)^2}.
\end{aligned}$$

So, in particular, the first equation above implies convergence of the R_n/n^2 as $n \to \infty$ whenever the sequence is sufficiently well behaved (see hypotheses below) since the sequence is Cauchy (for $m \ge n$):

$$\left|\frac{R_n}{n^2} - \frac{R_m}{m^2}\right| \le \left|\frac{S_n}{n} - \frac{S_m}{m} + \sum_{k=1}^{n-1} \frac{(n-k)}{n^2} X_k + \sum_{j=1}^{m-1} \frac{(m-j)}{m^2} X_j\right|$$

$$\le |\bar{S}_n - \bar{S}_m| + |\bar{S}_{m-1} - \bar{S}_n| + \left|\sum_{k=n}^{m-1} \frac{k}{n^2} X_k\right|$$

$$\le |\bar{S}_n - \bar{S}_m| + |\bar{S}_{m-1} - \bar{S}_n| + \left|\frac{m}{n^2}(S_{m-1} - S_n)\right|.$$

Don't you need absolute values here?

$\left|\sum_{k=n}^{m} kX_k\right| \le \sum_{k=n}^{m} k|X_k| \le m\sum_{k=n}^{m} |X_k|$ which is still not good enough

Now, by a rearrangement of the last expressions for R_n, we can write

$$R_n = (n-1)S_n - X_n - R_n + n \cdot X_n.$$

Thus by the SLLN we have that X_1 is integrable and $E[X_1] = m \iff$

$$\frac{2R_n}{n^2} \xrightarrow[n\to\infty]{a.s.} \lim_{n\to\infty} \left[\frac{(n-1)}{n}\bar{S}_n - \frac{X_n}{n^2} + \frac{X_n}{n}\right] = m.$$

This implies the claimed result. □

9.8 Problem 8

Problem Statement 9.9. Show that for a sequence of symmetric random variables $X_2, X_2, \ldots$ with finite moment generating function, and $S_n := \frac{2}{n(n+1)} \sum_{i=1}^n iX_i$, then

$$\mathbb{P}(|S_n| \ge x) \le 2\exp\left(-(n+1)I\left(\frac{nx}{n+1}\right)\right),$$

where

$$I(x) = \sup_{\lambda\in\mathbb{R}} \left\{\frac{x\lambda}{2} - \int_0^1 \Lambda(\lambda t)dt\right\},$$

where $\Lambda(\lambda) = \log M(\lambda)$ and $M(\lambda) = E[e^{\lambda X_1}]$. NOTE: Here, by a symmetric random variable X we mean that X and $-X$ have the same distributions.

Proof. See external solution PDF by I. Popescu using the standard tool for Large Deviations. □

9.9 Problem 9

Problem Statement 9.10. Find the rate function of the large deviation of a sequence of iid random variables with Poisson distribution with parameter $a < 0$.

Solution. If the $X_n \sim \text{Poisson}(a > 0)$, then for natural numbers $k \ge 0$, we have that

$$\mathbb{P}[X_n = k] = e^{-a}\frac{a^k}{k!}.$$

Since the sequence is iid, it suffices to find the rate function corresponding to X_1, which is defined to be

$$I(\varepsilon) = \sup_{\lambda\in\mathbb{R}} \left\{\varepsilon\lambda - \log E[e^{\lambda X_1}]\right\}.$$

Here, we have that the discrete expectation is defined by

$$E[e^{\lambda X_1}] = \sum_{k\ge 0} e^{\lambda k}\mathbb{P}[X_1 = k]$$

$$= \sum_{k\ge 0} \frac{(e^\lambda a)^k}{k!}$$

$$= e^{a(e^{\lambda}-1)}$$
$$\implies \log E[e^{\lambda X_1}] = a(e^{\lambda}-1).$$

Let the function we are trying to maximize within the rate function definition be defined by $g(\lambda) := \varepsilon\lambda - a(e^{\lambda}-1)$ for fixed ε. Then

$$g'(\lambda_0) = \varepsilon - ae^{\lambda_0} = 0$$
$$\implies \lambda_0 = \log\left(\frac{\varepsilon}{a}\right).$$

So the maximal values of this parameter corresponds to the rate function as

$$I(\varepsilon) = g(\lambda_0)\Big|_{\lambda_0=\log(\varepsilon/a)} = \varepsilon(\log(\varepsilon/a)-1)+a. \quad \square$$

9.10 Problem 10

Problem Statement 9.11. If $X_1, X_2, \ldots$ is a sequence of independent random variables so that $X_n \sim N(\mu_n, \sigma_n^2)$, then the following are equivalent:

1. $\sum_{n\geq 1} X_n$ converges almost surely;
2. $\sum_{n\geq 1}\mu_n$ and $\sum_{n\geq 1}\sigma_n^2$ are convergent series.

Proof of 1 $\implies$ 2. Let $S_n := \sum_{k=1}^n X_k$. Then as we know, $S_n \sim N(\mu_1+\cdots+\mu_n, \sigma_1^2+\cdots+\sigma_n^2)$. So it follows that if $S_n \xrightarrow{a.s.} X$ as $n\to\infty$, then what if $\sum \mu_n = +\infty$?

$$X \sim N\left(\sum_{n\geq 1}\mu_n, \sum_{n\geq 1}\sigma_n^2\right).$$

why? It could be that X is not really this normal.

Now if either of these sums diverges, or in the mean sum case fails to exist, then the limiting random variable X does not have a definite distribution to which S_n almost surely converges. So both must be convergent series. $\square$

Proof of 2 $\implies$ 1. Let $\mu := \sum_{n\geq 1}\mu_n$ and $\sigma^2 := \sum_{n\geq 1}\sigma_n^2$. Using the notation above we see that

$$\lim_{n\to\infty} S_n \sim N(\mu, \sigma^2) =: X,$$

which by assumption exists by the convergence of these sums. Moreover, by the convergence of these sums, given any $\varepsilon > 0$ we have that for all sufficiently large $n > N(\varepsilon)$, $\mathbb{P}(|S_n - X| > \varepsilon) < 2^{-n}$. This implies that

$$\sum_{n\geq 1}\mathbb{P}(|S_n - X| > \varepsilon) \leq \sum_{n=1}^{N(\varepsilon)}\mathbb{P}(|S_n - X| > \varepsilon) + \sum_{n>N(\varepsilon)} 2^{-n} < \infty.$$

So we conclude that S_n converges almost surely to the limiting X. $\square$

"$\Rightarrow$" If we take the characteristic function, then $E[e^{itS_n}] = e^{it\sum_{k=1}^{n}\mu_k - \frac{t^2}{2}\sum\sigma_i^2}$ which converges to something. This is enough to get that $\sum \mu_k$ & $\sum \sigma_i^2$ converge.

"$\Leftarrow$" We can use Kolmogorov's convergence theorem.

Problem 11

First if $X_n \Rightarrow X$ & $Y_n \Rightarrow Y$ then (X, Y indep)

$(X_n, Y_n) \Rightarrow (X, Y)$ since this can be seen

from $F_{X_n, Y_n}(x, y) = \mathbb{P}(X_n \le x, Y_n \le y) =$

$$= \mathbb{P}(X_n \le x)\,\mathbb{P}(Y_n \le y)$$

$$\xrightarrow[n\to\infty]{} \mathbb{P}(X \le x)\,\mathbb{P}(Y \le y)$$

$$= F_{X,Y}(x, y)$$

Similarly to the case of a single variable we can show then that

$$\mathbb{E}[\varphi(X_n, Y_n)] \to \mathbb{E}[\varphi(X, Y)]$$

for any continuous bounded $\varphi: \mathbb{R}^2 \to \mathbb{R}$.

In particular, for $f: \mathbb{R}^2 \to \mathbb{R}$ and any $\psi: \mathbb{R} \to \mathbb{R}$ cont and bounded,

$\varphi(x, y) = \psi(f(x, y))$ is cont and bounded on $\mathbb{R}^2$, so

$$\mathbb{E}[\varphi(X_n, Y_n)] = \mathbb{E}[\psi(f(X_n, Y_n))] \to \mathbb{E}[\psi(f(X, Y))]$$

In particular $f(X_n, Y_n) \Rightarrow f(X, Y)$.

9.11 Problem 11

Problem Statement 9.12. If X_n, Y_n are independent random variables such that $X_n \implies X$ and $Y_n \implies Y$, then for any continuous function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(X_n, Y_n) \implies f(X, Y)$ where X, Y are assumed to be independent.

Proof. Observe that since f is continouous, it is continuous in each input. Now since we have the convergence in distribution of X_n, Y_n this implies that for any $\varepsilon > 0$, $\exists N$ such that $n \geq N$ implies that each of

$$|E[f(X_n, Y_n)] - E[f(X_n, Y)]|, |E[f(X_n, Y)] - E[f(X, Y_n)]|, |E[f(X, Y_n)] - E[f(X, Y)]| < \varepsilon/3.$$

why?

Thus for any $n \geq N$ the triangle inequality implies that

$$|E[f(X_n, Y_n)] - E[f(X, Y)]| < \varepsilon.$$

This shows that $f(X_n, Y_n) \implies f(X, Y)$. This requires a little more care □

9.12 Problem 12

Problem Statement 9.13. Assume that $X_n \sim \text{Poisson}(n)$ and $Y_m \sim \text{Poisson}(m)$ are independent random variables. Show that

$$\frac{(X_n - n) - (Y_m - m)}{\sqrt{n + m}},$$

converges in distribution to $N(0, 1)$ when both $n, m \to \infty$.

Proof. The key here is in noticing that since X_n, Y_m are independent, $X_n + Y_m \sim \text{Poisson}(n + m)$, which has corresponding mean $\mu = n + m$. So the Central Limit Theorem immediately implies the result. To see why the sum of random variables is distributed this way, observe that by independence:

$$\begin{aligned}
\mathbb{P}(X_n + Y_m = k) &= \sum_{i=0}^{k} \mathbb{P}(X_n + Y_m = k, X_n = i) \\
&= \sum_{i=1}^{k} \mathbb{P}(Y_m = k - i, X_n = i) \\
&= \sum_{i=1}^{k} \mathbb{P}(Y_m = k - i)\, \mathbb{P}(X_n = i) \\
&= \sum_{i=1}^{k} \frac{e^{-m} m^{k-1}}{(k-i)!} \frac{e^{-n} n^i}{i!} \\
&= \frac{(m+n)^k}{k!} e^{-(m+n)},
\end{aligned}$$

which is also a Poisson distribution. This is ok. Now you have to use the fact that any Poisson(k) is a sum of k iid Poisson(1). □

9.13 Problem 13

Problem Statement 9.14. Take $Y_1, Y_2, \ldots$ to be iid random variables with $\mathbb{P}(Y_i = -1) = \mathbb{P}(Y_i = 1) = 1/2$, and construct $X_n := Y_n Y_{n+1}$. Find the variance of $S_n = X_1 + \cdots + X_n$ and show that

$$\lim_{n \to \infty} E[\hat{S}_n^k] = E[Y^k],$$

where $Y \sim N(0, 1)$ and $\hat{S}_n := S_n / \sqrt{\text{Var}(S_n)}$.

Proof. The first task is to compute the variance of S_n which we do by expanding the following expectations where we can easily compute that $E[Y_i^{2m+1}] = 0$ and $E[Y_i^{2m}] = 1$ for integers $m \geq 0$:

$$\begin{aligned} E[S_n] &= E\left[\sum_{i=1}^{n} Y_i Y_{i+1}\right] = \sum_{i=1}^{n} E[Y_i]^2 = 0 \\ E[S_n^2] &= E[S_{n-1}^2] + 2E[Y_n Y_{n+1}]E[S_{n-1}] + E[Y_n^2]^2, \quad \text{by independence,} \\ &= E[S_{n-1}^2] + 1 \implies E[S_n^2] = n. \end{aligned}$$

So we have that $\operatorname{Var}(S_n) = n$. Now for $k \geq 3$, we can similarly expand the expectations recursively to find that

$$E[S_n^k] = \sum_{i=0}^{k} \binom{k}{i} E[S_{n-1}^{k-i}]E[Y_1^i]^2 = E[S_{n-1}^k] + \sum_{i=1}^{\lfloor k/2 \rfloor} \binom{k}{2i} E[S_{n-1}^{k-2i}]. \tag{$*$}$$

So by induction with $(*)$, we can argue that $E[S_n^k] = 0$ whenever k is odd. Similarly, by an inductive argument, we can argue that when $k = 2m$ is even:

$$E[S_n^{2m}] = (2m-1)!! \cdot n^m + O(n^{m-1}).$$

We also know by results on Bernoulli polynomials (or Faulhaber's formula) that

$$\sum_{i=1}^{n} i^m = \frac{n^{m+1}}{m+1} + O(n^m).$$

Then using $(*)$ again in this case, we can see that the leading-order polynomial terms in n are given by

$$E[S_n^{2m+2}] = \binom{2m+2}{2}(2m-1)!! \cdot \frac{n^{m+1}}{m+1} + O(n^m) = (2m+1)!! \cdot n^{m+1} + O(n^m).$$

Then taking the limiting cases for the moments of $\hat{S}_n$, we obtain that

$$\lim_{n\to\infty} E[\hat{S}_n^{\,2m+2}] = (2m+1)!! = E[Y^{2m+2}].$$

Similarly, for the odd power moments we get that the limit is zero, which matches with the odd moments of the standard normal distribution as well. ☐

9.14 Problem 14

Problem Statement 9.15. Let $n \geq 1$ be a fixed natural number and let $X_1, \ldots, X_n$ be iid geometric random variables with parameter p, i.e., $\mathbb{P}(X_i = k) = p(1-p)^{k-1}$. Let $Y_p := X_1 + \cdots + X_n$. Show that when $p \to 0$, pY_p converges in distribution to a gamma distribution, and determine which.

Solution. By a counting argument, and/or induction on n, we find that for integers $k \geq 1$:

$$\mathbb{P}(Y_p = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}.$$

Then we can compute the expectation that

$$E[Y_p] = \sum_{k \geq n} k \binom{k-1}{n-1} p^n (1-p)^{k-n} = \frac{n}{p},$$

so that as p is a non-negative constant,

$$E[pY_p] = n.$$

Now we are trying to match this expectation to a gamma distribution G whose density function is given by

$$f_G(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \implies E[G] = \frac{\alpha}{n}.$$

So we can take $(\alpha, \lambda) := (1, 1/n)$ and obtain density function

$$f_G(x) = \frac{1}{n} e^{-x/n}.$$

TODO: How did / does this depend on the limiting p? □

This is not complete. The fact that $\mathbb{E}[pY_p] = n$ does not imply yet weak convergence!!

If we use characteristic functions, then

$f_X(t) = \mathbb{E}[e^{itX}]$, then

$$f_{pY_p}(t) = \mathbb{E}[e^{itpY_p}] = \prod_{k=1}^{n} \mathbb{E}[e^{itpX_k}] = \mathbb{E}[e^{itpX_1}]^n$$

Now $\mathbb{E}[e^{itpX_1}] = \sum_{l=1}^{\infty} e^{itpl} \cdot p(1-p)^{l-1}$

$$= \frac{p e^{itp}}{1-(1-p)e^{itp}}$$

$$= \frac{e^{itp}}{\frac{1-e^{itp}}{p} + e^{itp}}$$

$$\xrightarrow[p\to 0]{} \frac{1}{1-it}$$

which is the characteristic function of a geometric distribution with parameter 1.

Thus $f_{pY_p}(t) \xrightarrow[p\to 0]{} \frac{1}{(1-it)^n} = f_Z(t)$

where $Z \sim \text{Gamma}(n, 1)$ with density $\frac{1}{n!} x^{n-1} e^{-x} dx$

Another way of proving this is to show that pX_1^p converges to an exponential random variable. In particular,

$$F_{pX^p}(x) = 1 - P\left(X^p > \frac{x}{p}\right) = 1 - (1-p)^{\frac{x}{p}} \xrightarrow[p \to 0]{} 1 - e^{-x}$$

which is the dist of an exp.

Then a sum of n independent exp, is actually a Gamma dist. This can be seen directly by using induction.

10 Misc key facts, results, and other reminders

- If $(A_i)_{i\geq 1}$ are almost sure events, then $A = \cap_{i\geq 1} A_i$ is an almost sure event, i.e., if $\mathbb{P}(A_i) = 1$, then $\mathbb{P}\left(\cap_{i\geq 1} A_i\right) = 1$.
- If X is independent of itself, then X is almost surely constant. One consequence of this is that if X_i are independent and $a_1 X_1 + \cdots + a_n X_n = 7$, then X_i is constant almost surely, as X_1 is equal to an expression independent of itself.
- $\{\sup_n X_n > t\} = \bigcup_{n\to 1}\{X_n > t\}$.
- For $\vartheta \in \mathbb{R}$, $|e^{\imath\vartheta} - 1| \leq |\vartheta|$;
- **Variations of Inclusion-Exclusion:** For any two sets $E, F \subset \Omega$, we can write:

  (A) $\chi_{E\cup F} = \chi_E + \chi_F - \chi_{E\cap F}$; and
  (B) $\mathbb{P}(E\cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E\cap F)$,

  where $\mathbb{P}(E\cap F) = \mathbb{P}(E)\cdot\mathbb{P}(F)$ iff E, F are independent. Also, for any three sets A, B, C: $\mathbb{P}(A\cup B\cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A\cap B) - \mathbb{P}(A\cap C) - \mathbb{P}(B\cap C) + \mathbb{P}(A\cap B\cap C)$.
- Special sequences of sets constructions: $A_n := \{\omega : |X(\omega)| > n\}$;
- Special choices of functions: Take $f_n := n\cdot\chi_{[0,\frac{1}{n}]}$, or $f_n := \frac{1}{n}\cdot\chi_{[0,n]}$;
- $|a-b| = \frac{\max(a,b)-\min(a,b)}{2}$.
- $x + y = \min(x,y) + \max(x,y)$.
- $\chi_{[x,x+c]}(t) = \chi_{[t-c,t]}(x)$, which can be combined with an application of Fubini or Tonelli to swap the orders of integration in a double or multiple integral.
- Whenever see convergence in probability, try to extract a subsequence which converges almost surely.
- Whenever see (i) positive g_n; and (ii) $g_n \to g$ a.e., immediately try Fatou's lemma.
- If cannot prove convergence in probability directly, think proof by contradiction.
- Let Δ denote the *symmetric difference of sets*: $A\Delta B = (A\setminus B)\cup(B\setminus A)$.
- $E\setminus F = (E\setminus A)\dot\cup(A\setminus F)$.
- $|A\setminus E|_e < |A\Delta E|_e$.
- $f(x) = \sum_{j=1}^n \chi_{E_j}(x) \implies \int_0^1 f = \sum_{j=1}^n |E_j|$.
- $CF\setminus CE = E\setminus F$ and $E_k = F_k\cup(E_k\setminus F_k)$.
- Let $E_k^{(j)} := \{x\in E_k : j-1\leq |x| < j\} = (E_k\cap\{x : |x| < j\})\setminus\{x : |x| < j-1\}$. Then $E_k^{(j)}$ is *bounded* and measurable. May need to assume that $|E_k| < \infty$ for some k (the other case is easier).
- $E\setminus\cap_{m\geq 1} F_m = \cup_{m\geq 1} E\setminus F_n$
- If $|E| < \infty$, then $\chi_E(x) \in L^1(\mathbb{R})$.

- Continuous on $E \implies$ measurable on E.
- $\mu(\{x : |f_n - f| + |g_n - g| \geq \varepsilon\}) \leq \mu(\{x : |f_n - f| \geq \varepsilon/2\}) + \mu(\{x : |g_n - g| \geq \varepsilon/2\})$
- $\inf_n x_n \leq \liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n \leq \sup_n x_n$
  where $\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf_{m\geq n} x_m = \sup_{n\geq 0} \inf_{m\geq n} x_m$ and
  $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{m\geq n} x_m = \inf_{n\geq 0} \sup_{m\geq n} x_m$.
- $f(x) = \frac{1}{(x\log x)^{1/q}} \in L^q \notin L^p$ for $p < q$.
- Let $A_n := \{|f| > n\}$ and $A := \{|f| = \infty\}$. Then $A_{n+1} \subseteq A_n$ and $A = \int_{n\geq 1} A_n$ so that $A_n \searrow A$. By continuity, $\lim_{n\to\infty} |A_n| = |A| = 0$ when f is finite a.e.
- f' integrable $\implies$ for all $\varepsilon > 0$, $\exists \delta > 0$ such that for any measurable $A \subseteq \mathbb{R}$: $|A| < \delta \implies \int_A |f'| < \varepsilon$.
- $\liminf(s_n) + \liminf(t_n) \leq \liminf(s_n + t_n) \leq \limsup(s_n + t_n) \leq \limsup(s_n) + \limsup(t_n)$, for any sequences $\{s_n\}$ and $\{t_n\}$.
- $ab \leq \frac{1}{2}\left(\frac{a}{2} + b\right)^2$ and $ab \leq \frac{1}{4}(a+b)^2$.
- $1 + 1/2 + \cdots + 1/n - (1 + 1/n) < \log(n) < 1 + 1/2 + \cdots + 1/n$, so that $\frac{1+1/2+\cdots+1/n}{\log(n)} \to 1$ as $n \to \infty$.